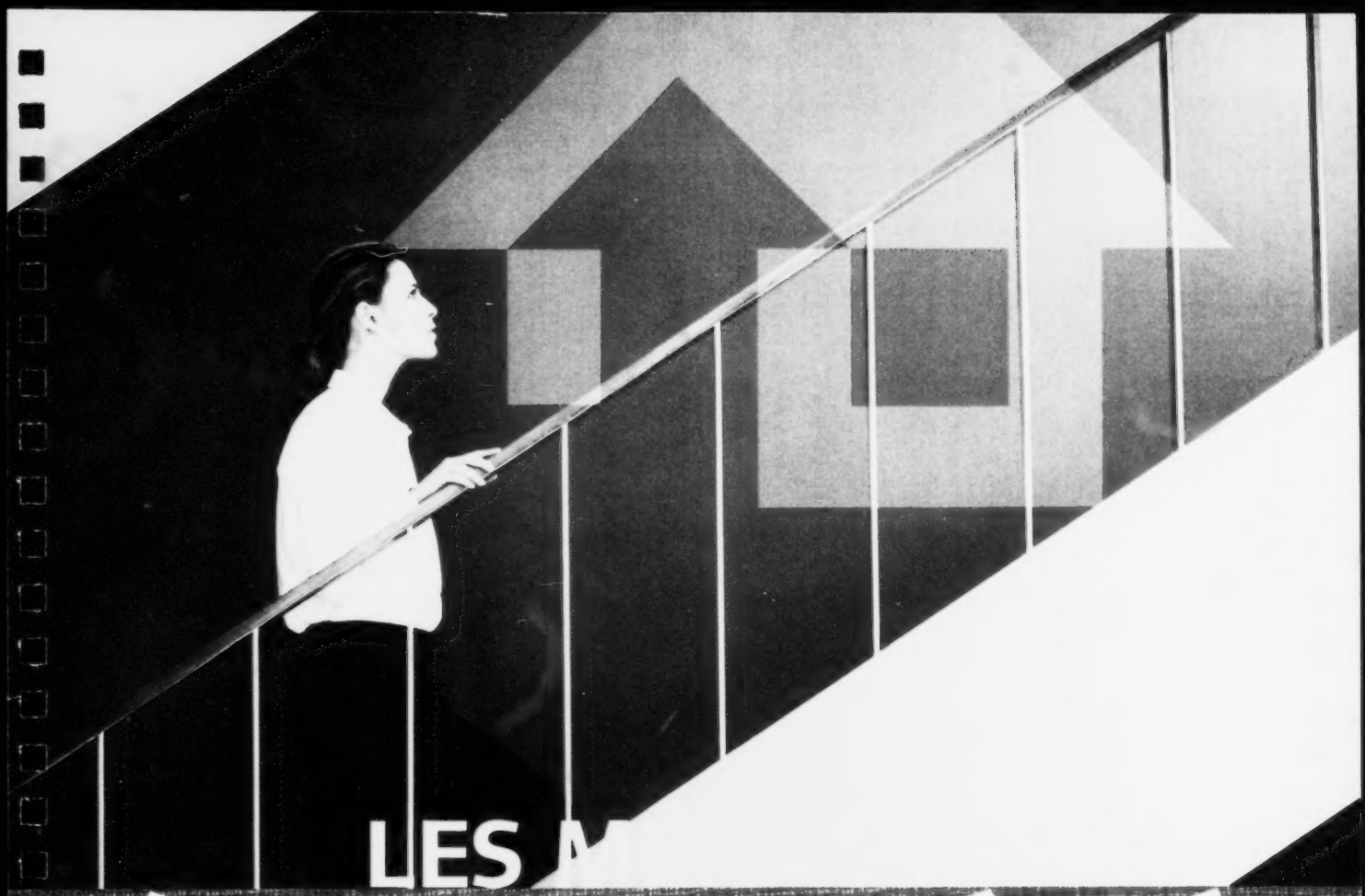

LES M

À TAUX D'EFFORT ÉLEVÉ

ANALYSE DE DONNÉES PROVENANT DE L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU (2002-2004)

Québec

LES MÉNAGES À TAUX D'EFFORT ÉLEVÉ

ANALYSE DE DONNÉES PROVENANT DE L'ENQUÊTE SUR LA DYNAMIQUE DU TRAVAIL ET DU REVENU (2002-2004)

AVRIL 2008

PAR FRANÇOIS RIVEST
SOCIÉTÉ D'HABITATION DU QUÉBEC

Le contenu de cette publication a été produit par François Rivest, de la Direction de la planification et du développement de programmes, en vertu de la programmation de recherche de la Société d'habitation du Québec.

Note : Les idées exprimées dans ce document ne traduisent pas nécessairement la position de la Société d'habitation du Québec; elles n'engagent que la responsabilité de l'auteur.

Avertissement

Dans les tableaux utilisant les pourcentages, la somme de ces derniers n'est pas nécessairement égale à 100 %, en raison de l'arrondissement des données.

Édition	Direction des communications
Graphisme de la couverture Mise en page	Direction des communications, Marlène Tremblay
Révision linguistique	Solange Deschênes
Impression et reliure	Copies de la Capitale

Société d'habitation du Québec
1054, rue Louis-Alexandre-Taschereau
Aile Saint-Amable, 3e étage
Québec (Québec) G1R 5E7

Toute reproduction totale ou partielle, par quelque procédé que ce soit, est interdite sans l'autorisation préalable écrite de la Société d'habitation du Québec.

Pour obtenir des exemplaires supplémentaires de ce document, on peut s'adresser au Centre de documentation de la Société d'habitation du Québec.

Succursale de Québec
Société d'habitation du Québec
Tél. : 418 643-4035 poste 1374

Succursale de Montréal
Société d'habitation du Québec
Tél. : 514 873-8775 poste 3081

Numéro sans frais : 1 800 463-4315
On peut également télécharger ou commander ce document à l'adresse Internet suivante :
www.habitation.gouv.qc.ca

Décembre 2008

Dépôt légal
Bibliothèque nationale du Québec, 2008
Bibliothèque nationale du Canada, 2008
ISBN 978-2-550-54673-3 (version imprimée)
ISBN 978-2-550-54674-0 (PDF)
© Gouvernement du Québec

NOTE DE TRANSMISSION AU DÉPOSITAIRE DES DONNÉES

Ce projet a été accepté en vertu du programme des Centres de données de recherche (CDR) de Statistique Canada. Plus précisément, il a fait l'objet de l'accord 06-POLICY-MTL-1006, conclu entre Statistique Canada, l'Institut de la statistique du Québec (coordonnateur des demandes en provenance des ministères et organismes de la fonction publique québécoise) et la Société d'habitation du Québec.

Cet accord autorisait l'utilisation des données confidentielles de l'*Enquête sur la dynamique du travail et du revenu* aux fins du projet de recherche soumis. Tous les calculs effectués à l'aide de ces données sont la responsabilité de l'auteur, tout comme le sont l'utilisation et l'interprétation de celles-ci.

Tel que prévu aux modalités de cet accord, le présent document a été acheminé à Statistique Canada, l'autorité responsable des CDR.

REMERCIEMENTS

Ce projet a été hébergé à l'antenne de l'Université Laval du Centre de données de recherche du Centre interuniversitaire québécois de statistiques sociales (CIQSS). L'auteur remercie le personnel du Centre pour son soutien, notamment M^me^ Danielle Forest (Montréal), qui a piloté une analyse longitudinale, et M.Charles Fleury, adjoint à la statistique à Québec.

Par ailleurs, ce projet a bénéficié, par contrat de services, de l'aide d'un professionnel de la Direction de la méthodologie et des enquêtes spéciales de l'Institut de la statistique du Québec, M. Maxime Boucher. L'auteur le remercie vivement de sa collaboration.

M. Jacques Trudel, de notre direction, a fait une lecture attentive du document, et plusieurs de ses suggestions ont été retenues. L'auteur remercie également M^mes^ Céline Giguère et Letsy Vernet pour leur collaboration à la mise en forme des tableaux et du texte.

TABLE DES MATIÈRES

Résumé **9**
Chapitre 1 : Introduction **13**
1.1 Objectif 13
1.2 Les données disponibles 13
1.2.1 Le recensement canadien 13
1.2.2 L'Enquête sur la dynamique du travail et du revenu 14
1.3 Les normes relatives au logement et à l'abordabilité 14
1.4 Le regroupement des taux d'effort 18
1.5 Le choix des variables indépendantes 20
Chapitre 2 : Analyses transversales **27**
2.1 Relations bivariées sous l'angle de la variable dépendante 27
2.2 Analyses multivariées transversales 30
2.3 Rapports de cotes des modèles multivariés transversaux 32
2.3.1 Rapports de cotes internes aux logits et aux variables 36
2.3.2 Rapports de cotes relatifs à la différenciation des ménages à fardeau moyennement élevé et à fardeau très élevé 37
Conclusion 39
Chapitre 3 : Premier regard sur les données longitudinales **41**
3.1 Le portrait des cheminements en matière de taux d'effort 41
3.2 Le modèle longitudinal avec fichier de mesures répétées 43
Conclusion 48
Chapitre 4 : Modèles longitudinaux avec dimension temps inhérente **51**
4.1 Modèle de taux d'effort élevé persistant 51
4.2 Modèle de taux d'effort élevé par trajectoire 55
Conclusion 61
Conclusion générale **63**
Bibliographie **70**

ANNEXES :
Annexe 1 Précisions relatives à la norme du rapport du coût d'habitation au revenu 73
Annexe 2 Comparaison des taux d'effort des ménages exploitant une ferme sur leur propriété avec celui des autres ménages, Québec, 2003, données pondérées complètes 75
Annexe 3 Taux d'effort des ménages selon leur composition, Québec, 2003, données pondérées 77
Annexe 4 Taux d'effort des ménages selon l'indicateur « Autres revenus », Québec, 2003, données pondérées 79
Annexe 5 Taux d'effort des ménages selon le panel de l'enquête auquel appartient le soutien économique principal, Québec, 2003, données pondérées 81
Annexe 6 Répartition des niveaux de revenu par principale source de revenus du ménage, fichier transversal, Québec, 2003, données pondérées 83
Annexe 7 Variables, niveaux et années pour lesquels le rapport de cotes du logit I est sensiblement plus accentué que celui du logit II, ou est seul significatif 85
Annexe 8 Extrait d'une présentation des modèles de trajectoires 87
Annexe 9 Taux d'effort annuel des ménages québécois, fichier longitudinal 2002-2004, données pondérées 89

LISTE DES TABLEAUX

Tableau 1.1 Taux d'effort détaillés, Québec, 2003, données brutes 17

Tableau 1.2 Ventilation des taux d'effort, Québec, 2003, données populationnelles (pondérées) 18

Tableau 1.3 Liste des principales variables de l'Enquête sur la dynamique du travail et du revenu utilisées 21

Tableau 1.4 Distribution des variables indépendantes selon les taux d'effort, Québec, 2003 23

Tableau 2.1 Taux d'effort selon les principales variables indépendantes, Québec, 2003 28

Tableau 2.2 Modèles transversaux, seuils observés des variables retenues pour les analyses transversales des fichiers de 2002, 2003 et 2004 31

Tableau 2.3 Fichiers transversaux 2002-2004 : rapports de cotes et seuils observés 34

Tableau 2.4 Variables, niveaux et années pour lesquels le rapport de cotes du logit I est non significatif, contrairement à celui du logit II 38

Tableau 3.1 Représentation < nombre et indice > des vecteurs de taux d'effort [n_1, n_2, n_3] 41

Tableau 3.2 Vecteurs des cheminements des ménages, Québec, 2002-2004, données pondérées 42

Tableau 3.3 Coefficients de corrélation des taux d'effort pour la période 2002-2004 (et seuils observés), Québec, données pondérées 43

Tableau 3.4 Modèle – Mesures répétées, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 45

Tableau 3.5 Modèle – Mesures répétées, seuils observés des interactions temps*variables indépendantes, fichier longitudinal 2002-2004 45

Tableau 3.6 Rapports de cotes, longitudinal 2002-2004 – Mesures répétées 46

Tableau 4.1 Modèle – Persistant, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 52

Tableau 4.2 Rapports de cotes, longitudinal « persistant » 53

Tableau 4.3 Score moyen de la variable dépendante pour les groupes issus de la fonction de mélange 56

Tableau 4.4 Modèle – Trajectoire, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 57

Tableau 4.5 Rapports de cotes, longitudinal 2002-2004 – Trajectoires 58

Tableau 4.6 Regroupement des vecteurs dans le modèle longitudinal par trajectoire 60

RÉSUMÉ

Cette étude sur les ménages à taux d'effort élevé est fondée exclusivement sur les données de l'Enquête sur la dynamique du travail et du revenu (EDTR), en fait, sur l'échantillon des ménages québécois de cette enquête pour la période de 2002 à 2004 (qui couvre la fin du panel 3 et le début du panel 4). L'EDTR est une enquête longitudinale pour laquelle les personnes retenues (identifiées à partir de l'Enquête sur la population active) sont suivies pendant six ans. La période de 2002 à 2004 a été retenue parce que les variables sur le logement nécessaires à cette étude ont été introduites dans l'EDTR en 2002, tandis que 2004 constituait la dernière année disponible (en 2006) et la fin d'un panel.

Le chapitre 1 expose les sources de données disponibles et les concepts fondamentaux de l'étude, tels que les trois normes associées à un logement acceptable (qualité convenable, taille convenable et coût abordable, c'est-à-dire coût d'habitation inférieur à 30 % du revenu avant impôt du ménage) et le fait que l'étude sera centrée sur le rapport du coût d'habitation au revenu du ménage, soit le taux d'effort ou le rapport d'abordabilité[1]. Les critères d'inclusion de la population visée sont précisés : ce sont les ménages québécois ayant un revenu égal ou supérieur à 6 500 $ et ayant des coûts d'habitation inférieurs à 100 000 $. La ventilation des taux d'effort de ces ménages pour l'année 2003 est donnée au tableau 1.2 : le pourcentage des ménages qui vivaient dans des logements abordables (moins de 30 % du revenu) est de 81,5 %.

Le premier chapitre poursuit en précisant que les taux d'effort, aux fins des analyses multivariées qui seront menées, ont été regroupés en trois niveaux : les rapports abordables (moins de 30 % du revenu), les rapports avec fardeau moyennement élevé (de 30 % à 49,9 %) et les rapports avec fardeau très élevé (50 % et plus). Finalement, le chapitre se termine par un survol des principales variables EDTR utilisées et par une première élimination de certaines d'entre elles. Les variables restantes font l'objet d'un tableau de distributions (bivariées) en fonction des taux d'effort.

Le chapitre 2 s'ouvre sur de nouvelles distributions bivariées, cette fois celle des taux d'effort selon les niveaux des variables indépendantes parcourues précédemment. Les signes apparents d'un effet des variables indépendantes sont relevés. L'étude vise une approche comparative de l'effet de ces variables.

Un modèle multivarié transversal mettant en présence treize variables pour expliquer les taux d'effort est testé sur les données de chacune des années de la période. D'une année à l'autre, les résultats peuvent varier. Cependant, il y a une cohérence d'ensemble si l'on adopte la règle que les résultats soient significatifs au moins deux années sur trois.

Sous cette condition, il y a huit variables significatives : âge du soutien principal, sexe (du soutien), principale source de revenu, genre de famille économique, activité principale, taille de la région de résidence, statut de déménagement et niveau d'études.

Il reste cinq variables non significatives : statut d'immigration (du soutien principal), état de santé, état de l'incapacité, nombre de personnes de 16 ans ou plus avec des gains et mode d'occupation. Ce résultat particulier, qui rejette l'hypothèse d'une différence significative entre propriétaires et locataires, est un constat propre au modèle mis de l'avant dans les analyses transversales. Il

1. Abordabilité est un néologisme auquel on aura recours à l'occasion. Il est accepté dans *Le grand dictionnaire terminologique* de l'Office québécois de la langue française, dans sa relation avec la propriété.

indique que la variable mode d'occupation, une fois pris en compte l'effet des variables significatives, n'ajoute rien à l'explication du taux d'effort.

Le chapitre 2 poursuit avec un premier examen de l'effet des variables significatives, et il essaie de jeter un peu de lumière sur ce qui peut différencier les ménages avec fardeau des coûts moyennement élevé et ceux avec fardeau des coûts très élevé. Ces thèmes seront repris ultérieurement.

Le chapitre 3 aborde les données longitudinales. Le parcours en matière de taux d'effort indique que les trois quarts des ménages de la population longitudinale (fichier pondéré) demeurent, au cours des trois années, avec des taux d'effort abordables (moins de 30 % du revenu). Le quart des ménages connaissent donc au moins une année de taux d'effort avec fardeau moyennement élevé (de 30 % à 49,9 %) ou avec fardeau très élevé (50 % et plus).

Un premier fichier longitudinal est créé, avec une observation par ménage et par année (introduction de la variable temps). Puisque l'introduction de la variable temps dans le modèle n'est pas significative, ni par elle-même ni en interaction avec d'autres variables, ce modèle peut être considéré comme un amalgame des modèles transversaux. Onze des variables des modèles transversaux sont retenues pour tester le modèle. Seules les variables statut d'immigration et état de santé actuel sont laissées de côté (deux des variables non significatives). Les résultats de ce premier modèle longitudinal sont cohérents avec ceux des modèles transversaux. Les huit variables significatives de ceux-ci sont maintenues. État de l'incapacité et mode d'occupation sont de nouveau non significatives. Par contre, la variable nombre de personnes avec des gains est faiblement significative, sans doute aidée par la taille échantillonnale de ce modèle longitudinal et par l'hypothèse de l'absence de corrélation longitudinale. Elle sera considérée comme non significative.

Les résultats par variable nous disent que :

- Tous les groupes d'âge, incluant les jeunes soutiens de ménage (16 à 29 ans), affichent des probabilités supérieures à celles du groupe des 65 ans et plus (groupe de référence) de connaître des rapports d'abordabilité avec fardeau moyennement élevé ou avec fardeau très élevé.
- Les hommes (variable sexe) affichent des probabilités inférieures de connaître de tels rapports.
- Les soutiens principaux qui ont pour principale source de revenu le travail autonome de même que ceux qui ont (à ce titre) les transferts gouvernementaux ou d'autres revenus affichent des probabilités nettement supérieures à celles du groupe de référence (pensions de retraite et autres placements) de connaître des rapports avec fardeau moyennement élevé ou très élevé.
- Les personnes seules et, dans une moindre mesure, les ménages monoparentaux affichent des probabilités supérieures à celles du groupe de référence de la variable genre de famille (les couples sans enfants) de connaître des rapports avec fardeau modéré ou élevé.
- En ce qui concerne la variable activité principale, un seul groupe se détache significativement du groupe de référence (gens à la retraite et cas indéterminés) : ce sont les soutiens dont l'activité principale est d'« aller à l'école » (étudier). Leurs probabilités de connaître des rapports avec fardeau sont nettement plus élevées.

- La variable taille de la région de résidence indique que les ménages des régions rurales et ceux des petites régions urbaines (moins de 100 000 habitants) affichent des probabilités plus faibles que ceux des régions urbaines de 500 000 habitants et plus (groupe de référence) de connaître des taux d'effort avec fardeau.
- La variable statut de déménagement indique que les familles économiques qui déménagent à la suite d'une séparation, ou pour toute autre raison, affichent des probabilités supérieures à celles des familles qui ne déménagent pas et des familles qui accueillent un nouveau membre de connaître des taux d'effort avec fardeau.
- Finalement, la variable du plus haut niveau d'études du soutien principal indique que les ménages qui ont un diplôme ou un certificat universitaire se démarquent significativement (plus faibles probabilités) du groupe de référence (les ménages avec moins qu'un diplôme d'études secondaires) et sont moins susceptibles de connaître une année de taux d'effort avec fardeau.

Le chapitre 4 est consacré à modéliser la persistance dans les taux d'effort modérés ou élevés. Nous avons vu que, sur une période de trois ans, le quart des ménages connaissent une ou plusieurs périodes de taux d'effort de 30 % ou plus (fardeau moyennement élevé ou très élevé). De ce nombre (le quart des ménages), la répartition est de 0,4 et 0,6 entre ceux qui n'ont connu qu'une période de taux avec fardeau et ceux qui en ont connu plus qu'une. Il est défini que ces derniers (plus qu'une période) représentent le groupe des persistants.

Les variables en fonction desquelles ils se distinguent des autres ménages sont les mêmes que précédemment, à deux exceptions près. Premièrement, la variable statut de déménagement devient non significative. Autrement dit, un déménagement dans l'année 1 (un) ne peut expliquer deux ou trois années de taux d'effort avec fardeau. Deuxièmement, la variable mode d'occupation, quant à elle, devient significative. Les ménages propriétaires affichent des probabilités inférieures de connaître un taux d'effort avec fardeau pendant deux ou trois années.

La variable activité principale reste significative, mais on observe un changement important. Ce n'est plus l'activité d'aller à l'école qui est significativement différente du niveau de référence (retraite, cas indéterminés), mais l'activité de travailler. Les soutiens principaux qui travaillent affichent des probabilités inférieures de connaître deux ou trois années de taux d'effort avec fardeau.

L'étude tente également de répondre à la question de savoir ce qui différencie les ménages à taux d'effort avec fardeau très élevé de ceux qui sont avec fardeau moyennement élevé. Les variables qui semblent avoir un effet plus généralisé par rapport à ce qui différencie les ménages à taux d'effort très élevé et ceux à taux d'effort moyennement élevé sont l'âge, la taille de la région de résidence et le statut de déménagement. Par effet plus généralisé, nous entendons un effet qui se manifeste dans des rapports de cotes plus accentués pour tous ou la plupart des niveaux de ces variables significatives. Les rapports de cotes qui comparent les ménages avec fardeau très élevé aux ménages à rapport abordable (logit I) vont s'écarter de façon tangible de rapports de cotes correspondants comparant les ménages avec fardeau moyennement élevé aux ménages à rapport abordable (logit II).

Deux autres variables ont un effet que nous qualifions de localisé, en ce sens que l'on n'observe qu'un seul niveau de ces variables pour lesquelles le rapport de cotes correspondant du logit I est accentué par rapport à son vis-à-vis du logit II. Ces variables sont le genre de famille économique et le statut d'activité. La première met en évidence le genre de famille des personnes seules, et la

seconde met en évidence les soutiens principaux qui vont à l'école. Seuls ces niveaux de ces variables se distinguent entre le logit I et le logit II.

CHAPITRE 1 : INTRODUCTION

1.1 OBJECTIF

L'objectif premier de cette étude est, essentiellement, d'en connaître davantage sur les ménages qui doivent consacrer une part importante (élevée) de leurs revenus à l'habitation. Les organismes fédéral et provinciaux responsables du secteur de l'habitation ont défini une norme, sur laquelle nous reviendrons, qui dit qu'un logement est abordable si le ménage y consacre moins de 30 % de son revenu avant impôt. Le projet de recherche fera sienne cette norme.

La population cible de l'étude sera donc (au minimum) d'étudier particulièrement les ménages qui consacrent 30 % et plus de leur revenu à l'habitation. Qu'est-ce qui caractérise ces ménages? De plus, dans quelle mesure les ménages qui consacrent une part élevée une année vivent-ils une situation temporaire, ou s'agit-il d'une situation relativement stable, qui a tendance à se répéter?

1.2 LES DONNÉES DISPONIBLES

1.2.1 Le recensement canadien

La source fondamentale pour obtenir des données sur le coût du logement pour les ménages provient du recensement du Canada, qui se tient de nos jours tous les cinq ans. Statistique Canada a prévu deux instruments de collecte pour la population des ménages. Il y a le formulaire 2A, qui est le formulaire simplifié destiné à la grande majorité des ménages, et il y a le formulaire 2B, le formulaire long, remis aléatoirement à environ un ménage sur cinq (20 % des ménages). C'est ce formulaire qui recueille de l'information sur le logement du ménage. Lors de l'année censitaire 1991, les questions sur le logement ont été élargies à la suite de demandes en ce sens de la part de la Société canadienne d'hypothèques et de logement (SCHL). Au recensement de 2006, ces questions étaient regroupées à la section H du formulaire long, qui tenait sur deux pages, et elles se répartissaient en huit questions principales. Les questions H6 à H8 permettent d'établir le coût du logement pour le ménage.

À l'intérieur du formulaire long, la section précédente au logement, soit la section G, recueille des données sur le revenu de chacun des membres du ménage. Cela rend possible de compiler le revenu total du ménage, et de mettre ce dernier en relation avec le coût du logement. Les bandes de données post-censitaires produites par Statistique Canada sur « le logement et les ménages » renferment, de façon agrégée, certains renseignements nécessaires à ces calculs. La Société d'habitation du Québec (SHQ), qui se procure ces bandes de données, est donc en mesure d'effectuer des relevés généraux du rapport entre le coût du logement et le revenu des ménages. Le terme qu'utilise fréquemment la SHQ pour désigner ce rapport est le taux d'effort. La SCHL préfère l'expression ratio des frais de logement au revenu du ménage (RFLR). Le terme rapport d'abordabilité est aussi employé.

Toutefois, pour étudier la question de la stabilité dans le temps du taux d'effort, les données du recensement constituent un outil mal adapté, à cause, premièrement, de l'intervalle de cinq ans entre les recensements et, aussi, parce que les ménages qui se voient remettre le questionnaire long varient d'un recensement à l'autre (leur distribution est aléatoire). Cette question de la stabilité invite donc à regarder auprès de bases de données dites longitudinales, c'est-à-dire des bases de données qui compilent des renseignements sur leurs unités échantillonnales de façon répétée sur une période de temps déterminée. Des bases de données longitudinales sont souvent maintenues

par des organismes spécialisés en statistiques, tels que Statistique Canada et l'Institut de la statistique du Québec (ISQ).

1.2.2 L'*Enquête sur la dynamique du travail et du revenu*

En 1993, Statistique Canada a lancé une enquête longitudinale dénommée l'***Enquête sur la dynamique du travail et du*** **revenu** (EDTR). Cette enquête suit deux panels d'environ 15 000 ménages sur une période de six ans chacun. Les panels sont décalés de trois ans (un nouveau panel tous les trois ans) et les membres du panel sont interviewés une fois l'an (période de collecte en janvier)[2]. Au tournant des années 2000, la SCHL a pressenti Statistique Canada pour qu'elle intègre à cette enquête une série de questions concernant le logement. Il s'agissait, pour la SCHL, d'une démarche double, puisqu'elle avait effectué une demande semblable relativement à l'Enquête sur les dépenses des ménages (EDM), une enquête annuelle transversale.

L'intégration d'informations sur le logement à l'intérieur de l'EDTR est devenue effective lors de l'année 2002, année qui coïncidait avec le renouvellement d'un panel. C'est donc dire qu'avec les années 2002 à 2004 nous disposons d'une période de trois ans pendant laquelle nous pouvons suivre l'évolution des coûts de logement pour les personnes échantillonnées par l'EDTR[3]. L'exploitation des données de cette enquête pour cette période devint l'objectif de ce projet de recherche. L'EDTR offre également aux chercheurs une version transversale (annuelle) de son fichier, que nous comptions aussi exploiter afin de déblayer le terrain et de tenter de distinguer ce qui serait propre aux analyses transversales et longitudinales du taux d'effort des ménages québécois[4].

1.3 LES NORMES RELATIVES AU LOGEMENT ET À L'ABORDABILITÉ

Il a été mentionné que la SCHL a été un partenaire de Statistique Canada pour intégrer de l'information sur le logement d'abord aux données du recensement et, plus tard, à l'Enquête sur les dépenses des ménages de même qu'à l'EDTR[5]. De plus, en collaboration avec les organismes provinciaux responsables de l'habitation, la SCHL a établi un ensemble de trois normes concernant le logement à l'aide desquelles elle examine les données produites à cet égard, en particulier, bien sûr, celles qui sont issues du recensement. Ainsi, après chaque recensement, la SCHL prépare une série de courtes publications dans la collection « Le point en recherche, série

2. Cette collecte concerne l'expérience des répondants sur le marché du travail au cours de l'année civile précédente (la dimension « travail »). Pour l'autre dimension de l'enquête, celle du revenu, environ 85 % des répondants autorisent l'utilisation des données fiscales. Les répondants qui ne fournissent pas cette autorisation font l'objet d'une deuxième collecte de données propre à cette dimension. Cette collecte se déroule en mai, immédiatement après la fin de la période des déclarations de revenu, lorsque les données sur le revenu sont fraîches à l'esprit des répondants.
3. Les données de 2004 sont devenues disponibles à l'été 2006. Il y a un intervalle de deux ans environ entre la collecte et la diffusion.
4. Le fichier transversal contient également un peu plus de sujets, parce qu'on lui joint un échantillon de rafraîchissement « pour garantir que l'échantillon total est représentatif de la population à un point donné dans le temps ». Statistique Canada, *Méthodologie de l'enquête sur la population active du Canada*, nº 71-526-XPB au catalogue, ministre de l'Industrie, 1998, p. 21.
5. Il est à noter que c'est la SCHL qui finance ces collectes de données. Voir SCHL, Le point en recherche, série socioéconomique, nº 3 révisé, *Qualité, taille et abordabilité du logement canadien*, avril 2004, note au tableau 1 p. 2.

socioéconomique », dans lesquelles elle traite, notamment, de l'évolution de la situation en regard à ces normes.

Un logement est dit acceptable s'il répond aux trois normes suivantes :

- Il est de qualité convenable, soit, de l'avis des occupants, il ne nécessite pas de réparations majeures.
- Il est de taille convenable, c'est-à-dire qu'il contient un nombre de chambres suffisant compte tenu de la taille et de la composition du ménage. Ce nombre de chambres suffisant est défini par la Norme nationale d'occupation[6].
- Il est de coût abordable, c'est-à-dire que les coûts d'habitation du ménage représentent moins de 30 % de son revenu avant impôt.

Cette recherche va s'intéresser uniquement à cette dernière norme. Celle-ci fait l'objet de précisions supplémentaires à l'annexe 1 (intitulée « Précisions relatives à la norme du rapport du coût d'habitation au revenu »).

L'intérêt de la SCHL pour ces normes n'est pas attribuable uniquement à des visées de recherche. Elles ont une application administrative bien concrète dans la gestion de certains programmes d'habitation puisque la SCHL a aussi mis au point le concept de ménages en besoins impérieux de logement. Un ménage éprouve un besoin impérieux si son logement ne répond pas à une ou plusieurs des trois normes susmentionnées, et s'il devrait consacrer 30 % ou plus de son revenu avant impôt pour payer le loyer médian exigible pour un logement répondant à ces trois normes dans sa localité de résidence. L'utilité de ce critère est d'éliminer des programmes d'aide les ménages qui ont les moyens de s'offrir un logement adéquat. Les ménages restants sont définis comme éprouvant des besoins impérieux et sont considérés prioritaires dans l'attribution de logements sociaux et de certains programmes d'aide au logement.

Cette recherche n'aborde pas le concept des besoins impérieux. Toutefois, nous verrons plus loin que nous avons prévu un critère pour éliminer certains ménages qui surconsomment au niveau de l'habitation.

Lorsqu'elle étudie les ratios des frais de logement au revenu avant impôt du ménage, la SCHL impose certaines restrictions à l'établissement de ces ratios.

- Elle ne considère pas les ménages agricoles et les ménages résidant dans les réserves indiennes.
- Elle ne considère pas les ménages dont le revenu est nul ou négatif.
- Elle ne considère pas les ménages dont le coût d'habitation est égal ou supérieur à leur revenu (rapports d'abordabilité de 100 % ou plus).

Ces restrictions, ou critères d'exclusion, éliminent environ 7 % des ménages (à l'échelle canadienne)[7]. Dans cette étude-ci, il a également fallu poser certains critères d'exclusion, mais en plus petit nombre.

---

6. Les prescriptions de cette norme figurent dans quelques-unes des publications de la SCHL de la série sur le logement selon les données du recensement. Voir, par exemple, SCHL 2004, *Qualité, taille et abordabilité du logement canadien*, *op. cit.*, page 1, note infrapaginale 1.
7. Voir SCHL, Le Point en recherche série socioéconomique nº 3 révisé, note infrapaginale 3, *Qualité, taille et abordabilité du logement canadien.*

La source de données en tant que telle, soit l'EDTR, ne couvre pas les ménages résidant sur les réserves indiennes. Il y a donc conformité de la base utilisée à cette balise de la SCHL. Par ailleurs, une autre balise de la SCHL exclut les ménages agricoles car la SCHL estime que ceux-ci ne peuvent distinguer les frais de possession de la maison familiale des dépenses associées aux structures agricoles de la ferme. Or la comparaison des taux d'effort des ménages agricoles québécois avec ceux des autres ménages à partir des données transversales de l'EDTR n'a pas montré de différence significative (voir annexe 2). En conséquence, les ménages exploitant une ferme sur leur propriété ont été retenus dans l'étude.

Pour les balises concernant le revenu et les coûts d'habitation, nous avons opté pour une approche différente. À partir du fichier de données transversales de l'année 2003 de l'EDTR, nous avons examiné la distribution de la variable dépendante, le taux d'effort, et de ses parties constituantes, le coût d'habitation et le revenu avant impôt, et avons résolu ce qui suit. Plutôt que d'éliminer les ménages dont le revenu est nul ou négatif, de même que ceux dont les coûts d'habitation sont égaux ou supérieurs à leurs revenus, nous avons commencé par éliminer tous les ménages dont le revenu est inférieur à 6 500 $.

Ce montant représente, *grosso modo*, l'équivalent de la prestation minimale annuelle d'aide sociale pour un ménage d'une personne seule en 2003 au Québec. Dans cette province, l'aide sociale est gérée par l'Administration provinciale, et l'admissibilité à ce programme est universelle[8].

Le raisonnement à l'appui de ce critère d'exclusion est le suivant : si des ménages échantillonnés par l'EDTR ne reçoivent pas l'aide sociale ou un montant équivalent, c'est qu'ils ont accès à des ressources ou à des actifs non rapportés par le questionnaire de l'EDTR[9]. Il s'agit donc d'une population possiblement différente, puisque le revenu de ces ménages, tel qu'il est mesuré par l'enquête, ne représente pas le niveau de précarité qu'il laisse supposer. Sinon, ces ménages recevraient de l'aide sociale, et leur taux d'effort en serait diminué d'autant.

Ce critère laissait de côté le problème du (relativement) petit nombre de ménages dont le taux d'effort était très élevé, parfois supérieur à 100 %, parce qu'ils dépensent des sommes très importantes pour se loger. En conséquence, les ménages dont le coût annuel d'habitation était supérieur à 100 000 $ ont été exclus. Ces observations étaient trop éloignées du reste de la distribution. Il s'agit d'une population de nature différente de celle qui nous intéresse. Nous voulons en savoir plus sur les ménages qui ont des taux d'effort élevés parce qu'ils ont des problèmes de faibles revenus, essentiellement. Cela ne peut guère être le cas des ménages qui s'engagent dans des frais d'habitation supérieurs à 100 000 $ par année[10].

Cette exclusion par rapport aux coûts d'habitation et ce seuil minimal sur les revenus ont été les moyens mis en œuvre pour éliminer des valeurs de taux d'effort qui paraissaient aberrantes, ou non conformes, au niveau de l'EDTR. Ces critères d'exclusion ont été laissés tels quels pour les

8. Notons que même les personnes temporairement sans adresse (ce qui exclut les personnes itinérantes) sont admissibles à ce programme. Voir Ministère de l'Emploi et de la Solidarité sociale, *Rapport statistique sur les programmes d'assistance sociale*, juillet 2007, tableau 1, p. 1-2.
9. Le ménage n'est pas interviewé sur les ressources et les actifs n'ayant pas à être déclarés comme des revenus : il peut s'agir d'économies ou de placements encaissables. Il peut également s'agir d'entrées en espèces telles que les dons en argent de personnes n'appartenant pas au ménage, d'héritages en espèces, de règlements d'assurance vie ou de gains nets provenant de jeux du hasard (voir Luffman, 2006, p. 23).
10. Lors de travaux plus récents, nous avons procédé à une vérification avec le *Fichier des microdonnées à grande diffusion du recensement de 2001* (fichier des ménages et du logement) et nous n'avons pas retrouvé de ménages québécois avec des coûts d'habitation supérieurs à 100 000 $ pour l'année 2000.

années 2002 et 2004 (en plus de 2003), de même que pour le fichier longitudinal[11]. Tous les enregistrements restants ont été considérés valides.

Le tableau qui suit présente la distribution des taux d'effort de l'échantillon transversal 2003 de l'EDTR tels qu'ils se présentaient avant les critères d'exclusion et après ceux-ci. Il s'agit des données brutes (le nombre d'enregistrements ménage).

Tableau 1.1 : Taux d'effort détaillés, Québec, 2003, données brutes

Taux d'effort	Tous les ménages		Ménages après exclusions		Variation	
	n	Cumulatif	n	Cumulatif	n	r
0 à 29,9 %	5 006	5 006	5 002	5 002	-4	-0,1 %
30 à 49,9 %	630	5 636	626	5 628	-4	-0,6 %
50 à 69,9 %	188	5 824	185	5 813	-3	-1,6 %
70 à 99,9 %	59	5 883	49	5 862	-10	-16,9 %
100 % et plus	74	5 957	22	5 884	-52	-70,3 %

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

On voit que l'imposition des critères d'exclusion réduit le bassin échantillonnal de 73 enregistrements, ou 1,86 % de ceux-ci[12]. Cinquante-deux de ces enregistrements relèvent de la catégorie des rapports de 100 % et plus[13], et dix, de la catégorie immédiatement inférieure, soit celle des 70 % à 99,9 %. Onze enregistrements seulement disparaissent dans les rapports entre 0 % et 69,9 % .

Nous croyons que, même si l'objectif est d'étudier les ménages aux taux d'effort élevés, il est préférable de faire preuve de discernement et de ne conserver que les enregistrements dont les données sur le revenu ont le plus de chances de représenter adéquatement la situation financière de ces ménages, du moins en contexte québécois. Si les revenus de certains ménages sont sous-estimés, le calcul du taux d'effort est tronqué pour ces cas. Il est à noter que l'EDTR dispose d'une variable qui indique la provenance des données sur le revenu pour le ménage[14]. Il s'agit le plus souvent de données fiscales de l'Agence du revenu (avec l'agrément du répondant)[15]. Sinon, ces données sont obtenues par entrevue, ou par imputation. Beaucoup de revenus très faibles (inférieurs à 2 000 $, par exemple) ont été obtenus par imputation.

11. Pour être précis, ces seuils n'ont été appliqués qu'à l'année de clôture (2004) en ce qui concerne le fichier longitudinal, afin de maximiser la taille de l'échantillon. La même règle prévaut pour le critère implicite de résidence au Québec.
12. Ce pourcentage ne se compare pas à celui qui a été établi par la SCHL (-7 %), non seulement parce que les données y sont non pondérées, mais aussi parce que le pourcentage de cette dernière tient compte des autochtones vivant sur les réserves et des ménages exploitants agricoles.
13. 70,3 % de cette catégorie est ainsi éliminée.
14. La variable INCSR42 (« source des informations sur le revenu »).
15. À l'échelle canadienne, « environ 85 % des répondants permettent d'utiliser les données fiscales plutôt que de répondre aux questions ». Willa Rea, et J. Yuen, *La dynamique de l'abordabilité des logements : résultats provisoires du projet de recherche conjoint SCHL-Statistique Canada*, présentation Powerpoint au Comité national de recherche sur le logement (CNRL), 1er mai 2007, diapositive 6.

Afin de mieux situer le lecteur, le tableau 1.2 présente les taux d'effort avec les données pondérées à la population des ménages, et ce, après la mise en place des seuils.

Tableau 1.2 Ventilation des taux d'effort, Québec, 2003, données populationnelles (pondérées)
Ménages après imposition des seuils (revenu ≥ 6 500 $ et coût habitation < 100 000 $)

	Taux d'effort (coût habitation / revenu avant impôt)			
Taux d'effort	**N** (000)	**%**	**N cumulatif** (000)	**% cumulatif**
0 à 29,9 %	2 518,0	81,5	2 518,0	81,5
30 à 49,9 %	380,1	12,3	2 898,1	93,9
50 à 69,9 %	132,5	4,3	3 030,6	98,1
70 à 99,9 %	44,6	1,4	3 075,2	99,6
100 % et plus	12,6	0,4	3 087,8	100,0

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

Sur la base des données de l'EDTR, ce tableau indique donc qu'à la suite de la mise en place des seuils, près de 82 % des ménages québécois avaient un taux d'effort respectant la norme en 2003 (moins de 30 % du revenu total va au logement). Au moyen des seuils élaborés par la SCHL, celle-ci avait établi, à partir des données du recensement de 2001, que 80,3 % des ménages québécois présentaient un taux d'effort à l'intérieur de la norme[16].

1.4 LE REGROUPEMENT DES TAUX D'EFFORT

En partant, on se retrouve donc avec un bassin de près d'un ménage sur cinq dont le taux d'effort dépasse la norme établie. On conviendra toutefois que les dépassements observés à cette norme sont moins préoccupants lorsqu'ils sont réduits, et que la préoccupation d'origine de cette étude concernait les taux d'effort plus élevés. Comme dans toute étude qui s'intéresse à un groupe particulier, la taille de la population visée devient une préoccupation importante lorsque celle-ci peut être définie de façon étroite.

Il faut, à cette étape-ci, rappeler que la SHQ a pu obtenir l'accès aux données confidentielles de l'EDTR en agréant à une entente de protection de la confidentialité des renseignements de ce fichier. Cette entente impose diverses restrictions, dont l'une porte sur la taille minimale des cellules d'un tableau de données. Cette entente ainsi que les besoins des techniques d'analyse utilisées font qu'il n'était pas possible de songer à créer un regroupement entre les rapports élevés de 70 % à 99,9 % et ceux de 100 % et plus. Comme l'indique le tableau 1.2, les deux ne totalisent pas tout à fait 2 % des ménages. L'étude aurait été grevée d'un manque de puissance et d'un potentiel analytique limité[17].

Par contre, en joignant les ménages occupant ces rapports supérieurs au groupe dont les rapports d'abordabilité vont de 50 % à 69,9 %, nous obtenons quelque 6 % des ménages, et cela peut

16. SCHL 2003, « Le point en recherche », série socio-économique, nº 1, *Amélioration sur le plan de l'abordabilité du logement*, donnée établie à partir du tableau 1, p. 4 . La différence de pourcentages entre ces deux sources de données indépendantes est relativement ténue.
17. Il aurait fallu limiter le nombre de variables et le nombre de catégories qu'elles comportent. L'utilisation des poids *bootstrap* pour l'estimation de la variance crée une contrainte supplémentaire.

constituer un groupe plus viable. Il faut donc envisager une étude des taux d'effort élevés sur le même modèle que celles de la SCHL et de J. Luffman (Statistique Canada). Parmi les publications « Le point en recherche », de la série socioéconomique, la SCHL a réalisé une étude distincte sur les « ménages qui consacrent 50 % ou plus de leur revenu au logement[18] ». C'est le seuil que fixe la SCHL pour les rapports d'abordabilité très élevés. Dans les autres publications de cette série, elle s'en tient généralement à la distinction entre les moins de 30 % et les 30 % et plus.

Une étude récente, de J. Luffman[19], reprend ces seuils et propose une terminologie particulière pour les désigner. Il y a les ménages sans fardeau des coûts d'habitation (taux d'effort inférieurs à 30 %), les ménages avec fardeau modéré (de 30 % à 49,9 %) et ceux avec fardeau élevé des coûts (de 50 % en montant). Ces deux derniers groupes constituent les ménages dont il est dit qu'ils éprouvent des problèmes d'abordabilité.

Cette étude-ci va également emprunter ces seuils et cette terminologie, en modifiant toutefois cette dernière légèrement : les fardeaux modérés deviendront les fardeaux moyennement élevés, et les fardeaux élevés deviendront les fardeaux très élevés. Ces deux groupes de rapports plus élevés, rappelons-le, ne font pas tout à fait 20 % des ménages. Les deux nous intéressent et nous essaierons de voir ce qu'il y a de particulier au niveau des 50 % et plus.

Dans les comparaisons qui seront menées, le groupe de référence de ces deux groupes aux taux d'effort élevés sera constitué des ménages dont les coûts d'habitation sont inférieurs à 30 %. En terminant, mentionnons que les données de la SCHL indiquent que lorsque le taux d'effort s'élève d'un fardeau moyennement élevé à un fardeau très élevé, la probabilité augmente que le ménage soit, selon l'expression de la SCHL, en « situation de besoins impérieux[20] ». À l'échelle canadienne, la probabilité d'être en situation de besoins impérieux est de 60,0 % parmi l'ensemble des ménages dont le fardeau des coûts égale ou dépasse 30 %, et cette probabilité s'élève à 81,6 % parmi les ménages dont le fardeau des coûts égale ou dépasse 50 %[21].

Le seuil du 30 % est donné par la norme de la SCHL et des organismes provinciaux d'habitation, norme dite d'abordabilité, et le seuil du 50 % est voulu par des considérations de taille échantillonnale, auxquelles s'ajoute le fait pratique de son utilisation par la SCHL[22] et par l'étude de Luffman. Le fait que cette étude partage les mêmes niveaux de taux d'effort est donc de nature à favoriser sa comparabilité avec ces autres études.

18. « Ménages qui consacrent 50 % et plus de leur revenu au logement ». Produit de la SCHL, nº 63481 (au catalogue SCHL). Paru dans la série socioéconomique 05-004, initialement en 2005, révisé en 2006 (SCHL, 2005).
19. J. Luffman, «Mesurer l'abordabilité du logement», revue *L'emploi et le revenu en perspective*, de Statistique Canada, nº 75-001-XIF au catalogue, nov. 2006, vol. 7, nº 11, p. 17-27.
20. Bien que cette étude ne traite pas des besoins impérieux de logement, rappelons que cette expression signifie que le logement du ménage ne répond pas à l'une ou plusieurs des trois normes (qualité, taille, abordabilité) et que ce ménage devrait débourser 30 % ou plus « de son revenu avant impôt pour payer le loyer médian exigé pour un logement local satisfaisant aux trois normes ». SCHL (2004), *Qualité, taille et abordabilité du logement canadien*, « Le point en recherche », série socioéconomique 04-007, p. 1 et 2.
21. SCHL (2005), *Ménages qui consacrent 50 % ou plus de leur revenu au logement*, « Le point en recherche », série socioéconomique 05-004. Pourcentages établis à partir des données du tableau 1, p. 2 (données applicables aux ménages privés non agricoles vivant hors réserve qui ont un revenu positif supérieur à leurs frais de logement).
22. SCHL (2005), *op. cit.*

1.5 LE CHOIX DES VARIABLES INDÉPENDANTES

Ce qui sera propre à l'un ou l'autre des niveaux de la variable dépendante (le taux d'effort ou rapport d'abordabilité) sera révélé par ce qu'il est convenu d'appeler des variables indépendantes. Celles-ci s'insèrent dans un modèle où leur rôle est d'expliquer la variation dans la variable dépendante.

En plus des études susmentionnées, le choix des variables indépendantes a été influencé par une première étude avec laquelle nous avions été mis en contact, et qui avait été faite à partir des données de l'EDTR. Il s'agit de l'étude de Bradley Brooks (de Statistique Canada), parue en 2005, et intitulée *Le chômage chronique : un profil statistique*[23]. Cette étude illustrait la grande diversité des variables présentes dans l'EDTR. Le thème n'est pas sans affinité avec le nôtre, puisque ce qui peut influencer une situation de chômage chronique peut également influencer une situation de taux d'effort élevés. Finalement, nous nous sommes inspiré également de notre lecture de la documentation de l'EDTR.

Comme nous l'avons mentionné précédemment, des analyses transversales ont d'abord été menées, et elles ont été suivies d'analyses longitudinales. La période d'expérimentation avec les variables indépendantes s'est déroulée au cours des analyses transversales. Certaines ont été essayées, pour être ensuite abandonnées. Parfois, un choix s'est effectué entre deux variables mesurant des concepts semblables. C'est aussi durant ces analyses que s'est établi le nombre de catégories ou de modalités des variables indépendantes. Des catégories peu ou moins populeuses ont pu être regroupées avec la catégorie de référence de la variable indépendante lorsque leur rapport de cotes[24] ne s'en distinguait pas significativement. Ces regroupements effectués au cours des analyses transversales ont généralement été conservés tels quels lors des analyses longitudinales.

Même si l'EDTR est un fichier constitué selon les personnes, le rapport d'abordabilité et les variables logement de l'EDTR sont des variables dont l'unité d'analyse est le ménage. Rappelons qu'un ménage est « une personne ou un groupe de personnes occupant un même logement[25] ». Conséquemment, la variable « indicateur de soutien économique principal du ménage » (MJIEH26) a été utilisée pour disposer des valeurs d'une personne en particulier du ménage lorsque la variable l'exigeait (la majorité d'entre elles l'exigent).

Les variables utilisées de façon exploratoire, et au long des analyses transversales, sont indiquées au tableau 1.3.

23. Produit de Statistique Canada, nº 11-621-MIF2005031, 14 pages (Brooks, 2005).
24. Rapport de cotes qui est en relation avec la variable dépendante. La définition d'un rapport de cotes sera abordée au prochain chapitre.
25. Statistique Canada (2002), *Dictionnaire du recensement de 2001*, p. 197.

Tableau 1.3 Liste des principales variables de l'*Enquête sur la dynamique du travail et du revenu* utilisées

Nom de code	Désignation officielle	Nom courant
age26	Âge de la personne au 31 décembre de l'année de référence (anréf)	Âge
crhlt26	État de santé actuel	
disabs26	Ind – État de l'incapacité pour l'anréf	Incapacité
dwtenr25	Logement possédé ou loué	Mode d'occupation
fmcomp27	Composition de la famille économique	Genre de famille
hhcomp25	Composition du ménage au 31 décembre de l'année de référence	Composit. du mén.
hlev2g18	Plus haut niveau d'instruction de la personne, rgp2	Niveau d'études
immst15	Indicateur – Personne est immigrante	
les26	Ind EVF* – Séparation dans la famille éco. durant l'anréf	Séparation
mjacg26	Activité principale de la personne durant l'anréf, rgp	
mjieh26	Indicateur (Ind) – Soutien économique principal du ménage	Soutien
mjsif27	Principale source de revenu pour la famille économique	
movst29	Statut de déménagement de la personne pour l'anréf	Déménagement
nbear27	Nbre de membres de la famille économique âgés 16+ avec gains	
opbu25	Ind – Exploitation d'entreprise à partir du mén./propriété	
opfm25	Ind – Exploitation d'une ferme sur cette propriété	
ottxm25	Total ménage – autre revenu	
panel21	Panel auquel appartient la personne	
rnbs25	Ind – Loyer payé est calculé en fonction de son revenu	
rnre25	Ind – Loyer mensuel subventionné par le gouvernement	
sex99	Sexe du répondant	Sexe
shely25	Coûts annuels d'habitation pour le ménage	
ttinc25	Total ménage – Revenu total avant impôt	
urbszg25	Taille de la région de résidence	
vismn15	Ind – Personne appartient à une minorité visible	

* EVF : Événement de la vie familiale

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu.

Puisqu'il faut, en analyse multivariée, éviter la redondance dans les variables indépendantes, certaines variables ont eu une vie utile de courte durée. C'est le cas, par exemple, de LES26, indicateur de séparation dans la famille économique. Non pas que la variable n'était pas significative, mais nous lui avons préféré la variable MOVST29 (statut de déménagement), dont l'une des modalités était « déménagé à la suite d'une séparation ».

De la même façon, VISMN15 – indicateur de minorité visible – était fortement relié à IMMST15 – indicateur d'immigrant. Cette dernière a été préférée en raison de son plus grand nombre de réponses positives.

La variable mode d'occupation (DWTENR25) est une variable importante de cette recherche. Elle crée une dichotomie entre ménages propriétaires et ménages non propriétaires (locataires). Depuis 2002, l'EDTR permet de raffiner les distinctions à l'intérieur du groupe des locataires à la suite de la création de deux variables indicatrices, soit RNBS25 – loyer payé en fonction du revenu – et RNRE25 – loyer subventionné. Étant donné le lien de ces deux variables avec les clientèles traditionnelles de la SHQ, elles ont été explorées, mais elles ont dû être laissées de côté, car elles créaient de trop petits domaines – sans compter le fait qu'elles étaient interreliées[26].

L'unité d'analyse

Il a été mentionné que l'unité d'analyse de la variable dépendante est le ménage. En effet, la variable dépendante crée le rapport de la variable SHELY25 et de la variable TTINC25, deux variables ménages, l'une étant divisée par l'autre. Cependant, les désignations officielles des variables EDTR utilisées (voir le tableau 1.3, colonne du centre) font souvent référence à d'autres unités, notamment celle de la famille économique (FMCOMP27, MJSIF27, NBEAR27, etc.). Ménages et familles économiques ne sont pas synonymes. Un ménage peut être composé d'une personne seule; il peut également compter plus d'une famille économique. Cependant, l'utilisation de la variable MJIEH26 nous procure l'ancrage du soutien économique principal (s.é.p.) du ménage, lequel, dans l'EDTR, inclut les personnes seules. Donc, même si un ménage compte plus d'une famille économique, MJIEH26 nous assure que c'est la situation de la famille économique ayant le plus haut revenu qui est considérée[27].

Les ménages qui comptent plus d'une famille économique ne sont pas nombreux. Ils étaient 232,4 dans l'échantillon en 2003 (le fichier transversal). Cela crée un petit domaine, d'exploitation difficile. De plus, le clivage important dans le tableau concernant la variable d'où est tirée cette information (variable Composition du ménage – voir l'annexe 3) est celui entre les ménages d'une personne et les ménages de deux personnes ou plus, formant une famille économique. Or, ce clivage est repris également par la variable genre de famille (FMCOMP27), puisque les ménages de personnes seules forment un des genres de famille définis par celle-ci. Cette dernière variable contenant plus d'informations, elle a donc été préférée dans l'analyse à celle de composition du ménage (HHCOMP25).

La variable du revenu total du ménage (TTINC25) va au dénominateur de la variable dépendante. Nous avons pensé initialement que ce rôle lui permettait d'être quand même utilisée comme variable indépendante. Elle fut donc insérée dans une première version du modèle explicatif en tant que variable catégorielle (et non pas continue). Après réflexion toutefois, tous les modèles transversaux ont dû être repris en excluant TTINC25. Ce retrait a pu jouer favorablement sur les variables restantes liées au revenu à l'intérieur du modèle.

La documentation de l'EDTR est parfois très succincte. C'était le cas notamment de la variable autre revenu (OTTXM25). Il n'était pas clair si les autres rentrées en espèces n'auraient pu être englobées sous ce chapeau. L'étude de Luffman semblait indiquer un lien positif entre ce type de

26. Par la suite, il a été porté à notre attention que l'EDTR permettait également de départager le groupe des propriétaires entre ceux qui détiennent une hypothèque et ceux qui n'en ont pas (à l'aide de la variable MORTG25). Le tableau 5 de l'étude de Luffman (2006, p. 26) montre cependant que la variation de rapport d'abordabilité entre ces deux groupes de propriétaires est beaucoup plus réduite qu'entre les ménages propriétaires (ensemble) et locataires.
27. C'est en vertu de cette caractéristique que MJIEH26 figure dans la liste des variables (tableau 1.3). Elle n'a pas été utilisée comme variable indépendante (explicative).

revenus et les taux d'effort (rapports d'abordabilité) chez les ménages locataires[28]. La vérification empirique avec les données de l'EDTR indique qu'il n'y a pas de relation (positive ou négative) entre les autres revenus et les taux d'effort globalement (propriétaires et locataires – voir l'annexe 4). Cela suggère que OTTXM25 ne recouvre pas les autres revenus en espèces[29].

Les ménages de l'EDTR sont suivis sur une période de six ans. C'est la durée d'un panel. Cette étude-ci inclut les ménages recrutés dans les panels 3 (années de fin de panel) et 4 (années de début de panel). L'étude de Brooks montrait un effet significatif de la variable panel (PANEL21), mais semblable effet ne s'est pas matérialisé dans cette étude-ci. Globalement, le panel auquel appartient un ménage n'a pas d'influence sur les taux d'effort durant la période 2002-2004 (voir l'annexe 5).

La plupart des variables restantes du tableau 1.3 ont nécessité des analyses plus diversifiées pour déterminer l'influence qu'elles ont sur le taux d'effort. Le prochain chapitre sera consacré aux analyses effectuées lors des analyses transversales. Auparavant toutefois, nous allons présenter les distributions des principales variables indépendantes de cette étude en fonction des rapports d'abordabilité (taux d'effort). Cela permet de mieux connaître l'échantillon, et cela constitue une première prise de contact avec les niveaux des variables indépendantes. L'évolution des pourcentages de chaque niveau donne également un certain aperçu de leur influence sur les taux d'effort. Notons que la variable niveaux de revenus est également incluse dans ce tableau. Elle est cependant placée en dernier, car elle ne sera pas utilisée dans les modèles explicatifs.

Tableau 1.4 Distribution des variables indépendantes selon les taux d'effort, Québec, 2003

Variables	Taux d'effort			Global
	Acceptable	Fardeau moy. élevé	Fardeau très élevé	
	%	%	%	%
Âge du soutien économique principal				
15 à 29 ans	12,0	13,7	17,6	12,5
30 à 44 ans	32,6	32,5	24,3	32,1
45 à 64 ans	37,9	25,9	40,4	36,5
65 ans et plus	17,6	28,0	17,8	18,9
Total	100,0	100,0	100,0	100,0
Sexe du soutien principal				
Homme	65,4	39,3	44,1	60,9
Femme	34,6	60,7	55,9	39,1
Total	100,0	100,0	100,0	100,0
Principale source de revenu de la famille économique				
Salaires et traitements	68,1	31,5	23,9	60,9
Revenu d'un travail autonome	3,4	6,5	11,5	4,3
Transferts gouvernementaux, autres revenus	18,0	58,4	59,9	25,6
Pensions de retraite, placements	10,5	3,7	4,7	9,3
Total	100,0	100,0	100,0	100,0

28. Autres rentrées en espèces plus élevées est en corrélation positive avec rapports d'abordabilité plus élevés. Voir le tableau 3 de l'étude de Luffman, aux colonnes Revenu.
29. Ceux-ci ne sont donc pas pris en considération par l'EDTR. Cet élément est discuté à l'annexe 1, de même qu'au début du prochain chapitre.

Tableau 1.4 Distribution des variables indépendantes selon les taux d'effort, Québec, 2003 (suite)

Variables	Taux d'effort			
	Acceptable	Fardeau moy. élevé	Fardeau très élevé	Global
	%	%	%	%
Genre de famille économique				
Personne seule	27,4	54,9	68,6	33,3
Couple sans enfants	26,9	14,4	9,4	24,3
Couple avec enfants	29,1	12,6	7,2	25,7
Famille monoparentale	5,7	14,0	9,9	7,0
Autres types	11,0	4,0	4,9	9,7
Total	100,0	100,0	100,0	100,0
Autres revenus du ménage				
Oui	18,8	18,7	15,2	18,6
Non	81,2	81,3	84,8	81,4
Total	100,0	100,0	100,0	100,0
Activité principale du soutien principal				
Travail	61,7	34,5	32,8	56,6
Étude	1,0	4,5	9,6	1,9
Retraite, codes réservés	28,8	34,7	25,4	29,3
Autres	8,6	26,3	32,1	12,2
Total	100,0	100,0	100,0	100,0
Taille de la région de résidence du ménage				
Rurale	20,3	14,0	13,8	19,1
Urbaine :				
Moins de 30 000 habitants	12,6	12,2	6,0	12,1
30 000 à 99 999 habitants	8,3	9,6	7,1	8,4
100 000 à 499 999 habitants	5,4	6,5	7,9	5,7
500 000 habitants et plus	53,4	57,7	65,3	54,7
Total	100,0	100,0	100,0	100,0
Statut d'immigration du soutien principal				
Immigrant	8,9	13,0	9,4	9,4
Non-immigrant	85,9	81,6	81,1	85,1
Codes réservés	5,2	5,4	9,5	5,5
Total	100,0	100,0	100,0	100,0
Statut de déménagement				
Déménagement du ménage	9,1	13,8	19,5	10,3
Déménagement après séparation	2,8	3,8	7,5	3,2
Ménage non déménagé, nouveau membre	88,2	82,4	73,0	86,5
Total	100,0	100,0	100,0	100,0
Niveau d'études du soutien principal				
Moins qu'un diplôme d'études secondaires	22,5	38,1	29,6	24,9
Diplôme d'études secondaires	21,5	21,4	19,0	21,3
Postsecondaire non universitaire	26,8	22,7	26,0	26,3
Diplôme ou certification universitaire	21,1	9,7	11,1	19,1
Codes réservés	8,1	8,2	14,4	8,5
Total	100,0	100,0	100,0	100,0

Tableau 1.4 Distribution des variables indépendantes selon les taux d'effort, Québec, 2003 (suite)

Variables	Taux d'effort			
	Acceptable	Fardeau moy. élevé	Fardeau très élevé	Global
	%	%	%	%
État de santé du soutien principal				
Excellent	24,1	19,3	14,8	22,9
Très bien	32,6	21,2	25,3	30,8
Bien	23,7	25,0	24,4	23,9
Passable	7,4	17,1	16,0	9,1
Mauvais	2,0	6,0	8,3	2,9
Codes réservés	10,1	11,3	11,0	10,3
Total	100,0	100,0	100,0	100,0
État de l'incapacité du soutien principal				
Incapacité	21,5	30,3	35,1	23,4
Non-incapacité	68,1	57,5	53,3	65,9
Codes réservés	10,4	12,2	11,7	10,7
Total	100,0	100,0	100,0	100,0
Mode d'occupation du ménage				
Propriétaire	65,4	36,2	35,0	60,0
Locataire	34,6	63,8	65,1	40,0
Total	100,0	100,0	100,0	100,0
Nombre de membres de la famille économique avec gains				
Aucun membre	17,4	42,0	47,0	22,3
Un membre	35,7	41,5	41,4	36,7
Deux membres ou plus	46,9	16,6	11,6	41,0
Total	100,0	100,0	100,0	100,0
Niveaux de revenu total du ménage				
6 500 à 14 999 $	2,8	29,8	66,7	10,0
15 000 à 24 999 $	9,0	39,9	19,8	13,5
25 000 à 34 999 $	15,4	15,4	7,7	14,9
35 000 $ et plus	72,9	14,9	5,9	61,6
Total	100,0	100,0	100,0	100,0
n	**4 767,6**	**719,7**	**359,0**	**5 846,3**
%	81,5	12,3	6,1	100,0

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

CHAPITRE 2 : ANALYSES TRANSVERSALES

INTRODUCTION

Ce chapitre rapporte les analyses transversales pour les années 2002, 2003 et 2004. Celles-ci ne peuvent pas nous renseigner sur la stabilité des taux d'effort (rapports d'abordabilité) pour les ménages eux-mêmes, mais peuvent nous renseigner sur la stabilité de la significativité ou non des variables indépendantes.

2.1 RELATIONS BIVARIÉES SOUS L'ANGLE DE LA VARIABLE DÉPENDANTE

Précédemment, le tableau 1.4 présentait les distributions propres aux variables indépendantes en fonction des trois niveaux de taux d'effort. L'objectif poursuivi était descriptif. Il voulait montrer comment se répartissaient les niveaux des variables indépendantes selon les trois niveaux de rapports d'abordabilité.

Le premier tableau de ce chapitre-ci reprend le format du tableau 1.4, mais tente d'explorer davantage le potentiel explicatif des variables indépendantes en réorganisant l'information de telle sorte que les proportions (exprimées sous forme de pourcentages) portent sur la distribution des taux d'effort selon les niveaux des variables indépendantes. Ainsi, la comparaison des proportions des niveaux d'une variable indépendante qui se retrouvent sous la colonne « Taux d'effort acceptable », par exemple, instruit davantage relativement au pouvoir explicatif de cette variable que la méthode utilisée au tableau 1.4. Plus il y a de variation dans les proportions sous une colonne donnée, meilleures sont les chances que cette variable influence significativement la variable dépendante (du moins, en relation bivariée et en supposant l'absence de cellules de taille insuffisante).

Le tableau 2.1 reprend, en fin de compte, la même façon de faire que celle qui est illustrée par les tableaux des annexes 2 à 5 (annexes liées au chapitre précédent). Il présente les distributions des taux d'effort selon les principales variables indépendantes de cette étude. À titre indicatif, le seuil observé du test du khi-deux, tel qu'il est mesuré par SUDAAN[30], de ces relations bivariées est également rapporté. Parmi ce groupe de quinze variables, ces seuils sont tous significatifs à $p < 0,01$, sauf celui pour autres revenus du ménage[31] (seuil de 0,602) et celui pour la variable statut d'immigration (seuil de 0,131). Celui pour la variable âge est sur la frontière (seuil de 0,059). Notons que la variable niveaux de revenus est présente, comme elle l'était au tableau 1.4, et livrée une fois de plus en dernier, puisqu'elle n'est pas utilisée dans les modèles explicatifs (elle constitue le dénominateur de la variable dépendante).

30. SUDAAN est un logiciel spécialisé de statistiques qui permet, notamment, de prendre en compte les plans de sondage complexes, comme celui de l'EDTR. Tous les seuils de signification de cette étude ont été calculés à l'aide des mille poids *bootstrap* développés par Statistique Canada pour cette enquête (utilisés avec la méthode dite des répliques répétées équilibrées). Notons que depuis sa version 8, SAS s'est aussi engagé dans cette voie (prise en compte des plans complexes), mais pas avec la méthode des répliques.
31. La présence de la variable autres revenus s'expliquait par notre désir de clarifier ce que recouvre cette variable dans l'EDTR (voir la discussion de l'annexe 1). Le fait qu'elle y soit non significative, alors qu'elle était significative dans l'étude de Luffman (voir le tableau 4, p. 24, de cette étude), nous indique qu'il s'agit des autres revenus imposables. Statistique Canada a éventuellement corroboré cette interprétation.

Tableau 2.1 Taux d'effort selon les principales variables indépendantes, Québec, 2003

Variables	Taux d'effort			Total	
	Acceptable	Fardeau moy. élevé	Fardeau très élevé		
	%	%	%	%	n
Âge du soutien économique principal					
15 à 29 ans	78,0	13,4	8,6	100,0	732,0
30 à 44 ans	82,9	12,5	4,6	100,0	1 873,9
45 à 64 ans	84,5	8,7	6,8	100,0	2 136,3
65 ans et plus	76,1	18,2	5,7	100,0	1 104,1
p. χ^2 SUDAAN = 0,059					
Sexe du soutien principal					
Homme	87,6	7,9	4,5	100,0	3 560,5
Femme	72,1	19,1	8,8	100,0	2 285,8
p. χ^2 SUDAAN = 0,000					
Principale source de revenu de la famille économique					
Salaires et traitements	91,2	6,4	2,4	100,0	3 557,4
Revenu d'un travail autonome	64,6	18,8	16,6	100,0	248,6
Transferts gouvernementaux, autres revenus	57,5	28,1	14,4	100,0	1 494,3
Pensions de retraite, placements	92,1	4,8	3,1	100,0	545,9
p. χ^2 SUDAAN = 0,000					
Genre de famille économique					
Personne seule	67,0	20,3	12,7	100,0	1 946,1
Couple sans enfants	90,3	7,3	2,4	100,0	1 419,5
Couple avec enfants	92,3	6,0	1,7	100,0	1 501,6
Famille monoparentale	66,8	24,6	8,6	100,0	410,0
Autres types	91,8	5,1	3,1	100,0	569,2
p. χ^2 SUDAAN = 0,000					
Autres revenus du ménage					
Oui	82,6	12,4	5,0	100,0	1 086,3
Non	81,3	12,3	6,4	100,0	4 760,1
p. χ^2 SUDAAN = 0,602					
Activité principale du soutien principal					
Travail	88,9	7,5	3,6	100,0	3 306,9
Étude	40,6	28,6	30,8	100,0	112,3
Retraite, codes réservés	80,1	14,6	5,3	100,0	1 714,8
Autres	57,3	26,5	16,2	100,0	712,3
p. χ^2 SUDAAN = 0,000					
Taille de la région de résidence du ménage					
Rurale	86,6	9,0	4,4	100,0	1 118,3
Urbaine :					
Moins de 30 000 habitants	84,7	12,3	3,0	100,0	709,7
30 000 à 99 999 habitants	80,7	14,1	5,2	100,0	490,0
100 000 à 499 999 habitants	77,3	14,1	8,6	100,0	331,2
500 000 habitants et plus	79,7	13,0	7,3	100,0	3 197,0
p. χ^2 SUDAAN = 0,0003					
Statut d'immigration du soutien principal					
Immigrant	76,9	17,0	6,1	100,0	551,3
Non immigrant	82,3	11,8	5,9	100,0	4 972,3
Codes réservés	77,4	12,0	10,6	100,0	322,7
p. χ^2 SUDAAN = 0,131					

Tableau 2.1 Taux d'effort selon les principales variables indépendantes, Québec, 2003 (suite)

Variables	Taux d'effort			Total	
	Acceptable	Fardeau moy. élevé	Fardeau très élevé		
	%	%	%	%	n
Statut de déménagement					
Déménagement du ménage	71,9	16,5	11,6	100,0	601,1
Déménagement après séparation	71,0	14,6	14,4	100,0	187,4
Ménage non déménagé, nouveau membre	83,1	11,7	5,2	100,0	5 057,8
p. χ^2 SUDAAN = 0,000					
Niveau d'études du soutien principal					
Moins qu'un diplôme d'études secondaires	73,8	18,9	7,3	100,0	1 454,6
Diplôme d'études secondaires	82,3	12,3	5,4	100,0	1 245,2
Postsecondaire non universitaire	83,4	10,6	6,0	100,0	1 535,7
Diplôme ou certification universitaire	90,3	6,2	3,5	100,0	1 113,6
Codes réservés	77,7	11,9	10,4	100,0	497,1
p. χ^2 SUDAAN = 0,000					
État de santé du soutien principal					
Excellent	85,6	10,4	4,0	100,0	1 338,7
Très bien	86,4	8,5	5,1	100,0	1 799,0
Bien	80,8	12,9	6,3	100,0	1 399,5
Passable	66,1	23,1	10,8	100,0	534,2
Mauvais	57,2	25,2	17,6	100,0	170,4
Codes réservés	79,9	13,5	6,6	100,0	604,4
p. χ^2 SUDAAN = 0,000					
État de l'incapacité du soutien principal					
Incapacité	82,4	7,5	10,1	100,0	1 245,7
Non-incapacité	80,6	14,6	4,8	100,0	4 024,2
Codes réservés	86,0	6,7	7,3	100,0	576,4
p. χ^2 SUDAAN = 0,0004					
Mode d'occupation du ménage					
Propriétaire	89,1	7,4	3,5	100,0	3 505,2
Locataire	70,4	19,6	10,0	100,0	2 341,0
p. χ^2 SUDAAN = 0,000					
Nombre de membres de la famille économique avec gains					
Aucun membre	63,8	23,2	13,0	100,0	1 302,3
Un membre	79,2	13,9	6,9	100,0	2 147,1
Deux membres ou plus	93,3	5,0	1,7	100,0	2 397,0
p. χ^2 SUDAAN = 0,000					
Niveaux de revenu total du ménage					
6 500 à 14 999 $	22,4	36,7	40,9	100,0	585,3
15 000 à 24 999 $	54,5	36,5	9,0	100,0	787,2
25 000 à 34 999 $	84,1	12,7	3,2	100,0	870,8
35 000 à 44 999 $	91,8	7,2	1,0	100,0	731,5
45 000 $ et plus	97,6	1,9	0,5	100,0	2 871,5
p. χ^2 SUDAAN = 0,000					
n	**4 767,6**	**719,7**	**359,0**	**5 846,3**	
%	**81,5**	**12,3**	**6,1**	**100,0**	

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

Même si elle ne peut être utilisée pour modéliser le rapport d'abordabilité, on constate aisément que la dernière variable du tableau, soit la variable niveaux de revenu, est la variable dont les variations de proportions à la colonne Acceptable sont les plus prononcées, allant de 22,4 % seulement des ménages ayant un revenu de 6 500 $ à 14 999 $ qui ont un rapport d'abordabilité acceptable, à 97,6 % pour ceux dont le revenu est égal ou dépasse 45 000 $.

La variable principale source de revenu de la famille économique (troisième en ordre de liste) est reliée à celle du niveau de revenu[32], et on y relève également d'assez bons écarts de proportions. Les ménages ayant les transferts gouvernementaux ou les autres revenus comme principale source n'ont un rapport d'abordabilité acceptable qu'à 57,5 %, alors que ceux qui ont les salaires et traitements comme principale source (qui forment la majorité des ménages) ont un rapport acceptable à 91,2 %. Ce contraste est très informatif.

La variable genre de famille économique montre aussi un contraste intéressant entre les familles de personnes seules (67,0 % acceptable), les familles monoparentales (66,8 % acceptable) et les autres familles (plus de 90 % acceptable). D'autres variables laissent voir un contraste intéressant au niveau des pourcentages de rapports d'abordabilité acceptables, entre autres l'activité principale du soutien principal (ses pourcentages vont de 88,9 % pour les soutiens qui travaillent, à 40,6 % pour ceux qui vont à l'école). S'y ajoutent aussi le niveau d'études (de 73,8 % à 90,3 %), l'état de santé (de 57,2 % à 86,4 %), le mode d'occupation (de 70,4 % à 89,1 %) et le nombre de membres de la famille économique de 16 ans ou plus avec gains (de 63,8 % à 93,3 %)[33].

2.2 ANALYSES MULTIVARIÉES TRANSVERSALES

Bien qu'elles soient intéressantes, ces remarques et observations ne procurent pas une façon éprouvée de discerner lesquelles parmi ces variables sont vraiment importantes pour expliquer les variations entre ménages dans les taux d'effort, une fois établi que la relation bivariée est significative. Avec des données pour lesquelles les variables indépendantes sont interreliées, il s'impose d'utiliser une technique d'analyse multivariée pour faire la part des choses. Ici, la technique de l'analyse de régression logistique multinomiale a été utilisée. Cette technique crée des équations ou des fonctions logits pour chaque niveau de la variable dépendante par rapport à son niveau de référence[34]. Comme elle produit un vecteur de paramètres distincts pour chacune des équations logits, elle utilise plus de degrés de liberté que la régression logistique ordinale à cotes proportionnelles (en particulier) et, conséquemment, penche du côté conservateur[35].

La régression logistique

La régression logistique établit le rapport entre la cote (probabilité) estimée qu'un événement ait lieu (taux d'effort de 50 % et plus, tiré de l'une des fonctions logits) parmi un groupe donné

32. L'annexe 6 illustre cette relation pour le fichier transversal de 2003.
33. Les variables les moins intéressantes sous ce rapport sont les variables dont le khi-deux n'est pas significatif, soit autres revenus et statut d'immigration.
34. La variable dépendante (le taux d'effort) comporte trois niveaux. La technique crée donc deux équations logits.
35. Bien que la variable dépendante de l'étude soit de nature ordinale (le taux d'effort), la régression logistique ordinale à cotes proportionnelles, qui est le modèle ordinal par défaut (SUDAAN et SAS), n'a pas été utilisée parce que les données ne se conformaient généralement pas au postulat qui sous-tend les rapports de cotes proportionnels. Voir à ce sujet Marc J. Gameroff, *Using the Proportional Odds Model for Health-Related Outcomes: Why, When, and How with various SAS Procedures*, Paper 205-30, SUGI 30, 8 p.

(sexe = hommes, par exemple), par rapport à la cote estimée qu'il ait lieu au sein d'un groupe de référence (sexe = femmes), en tenant constantes toutes les autres variables explicatives (ou indépendantes). Dans cet exemple, un rapport de cotes supérieur à 1,0 indique que l'événement en question a plus de chances de se produire chez les hommes que chez les femmes, et l'inverse pour un rapport de cotes compris entre 0,0 et <1,0. Les coefficients bêtas du modèle de régression logistique de ces rapports de cotes ont un seuil observé (une valeur-p) obtenu généralement par des tests de khi-deux[36].

Le tableau 2.3 reproduira les rapports de cotes et les seuils observés des coefficients bêtas pour les variables incluses dans les modèles transversaux des années 2002, 2003 et 2004. Auparavant, nous allons soumettre le tableau d'ensemble à l'égard des seuils observés des variables indépendantes que produit la procédure de régression logistique (SUDAAN)[37].

Tableau 2.2 Modèles transversaux, seuils observés des variables retenues pour les analyses transversales des fichiers de 2002, 2003 et 2004 (non significatif : seuil observé ≥ 0,05)

	Années		
Variables	**2002**	**2003**	**2004**
Âge	0,0000	0,0000	0,0003
Sexe	0,0005	0,0001	Non significatif
Princ. source revenu	0,0000*	0,0000	0,0000
Genre de famille	0,0000**	0,0000	0,0000
Activité principale	0,0047	0,0009	0,0000
Taille région résidence	0,0014	0,0000	0,0000
Immigration	0,0460	Non significatif	Non significatif
Déménagement	Non significatif	0,0009	0,0311
Niveau d'études	Non significatif	0,0037	0,0356
Santé	Non significatif	Non significatif	Non significatif
Incapacité	Non significatif	Non significatif	Non significatif
Mode occupation	Non significatif	Non significatif	Non significatif
Nbre pers. avec gains	Non significatif	Non significatif	Non significatif

* En 2002, la variable Principale source de revenu (Mjsif27) a été utilisée avec trois catégories plutôt que quatre (protection confidentialité).

** En 2002, la variable Genre de famille (Fmcomp27) a été utilisée avec quatre catégories plutôt que cinq (protection confidentialité).

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichiers transversaux 2002-2004.
Compilation : Société d'habitation du Québec.

La première chose qu'enseigne l'examen des résultats des modèles transversaux est qu'en matière de significativité les résultats ne sont pas toujours identiques d'une année à l'autre. Les fichiers transversaux de cette enquête longitudinale comptent suffisamment d'observations

36. Le test ajusté du khi-deux avec la correction de Satterthwaite aux degrés de liberté a généralement été utilisé.
37. Ces seuils observés sont produits à la suite de l'application de modèles de régression logistique multinomiale avec le logiciel SUDAAN. Comme nous l'avons mentionné précédemment, celui-ci permet de tenir compte du plan d'échantillonnage complexe de l'EDTR.

distinctes (une année par rapport à l'autre) pour engendrer une certaine variabilité des résultats[38]. Cependant, la majorité des variables obtiennent des résultats cohérents au cours des trois années.

Il y a le groupe des variables pour lesquelles les résultats sont significatifs les trois années : âge, principale source de revenus (MJSIF27), genre de famille économique (FMCOMP27), activité principale du répondant principal (MJACG27) et taille de la région de résidence (URBSZG25). Donc cinq variables sur treize.

Il y a le groupe des variables **non** significatives les trois années. Elles sont au nombre de quatre : nombre de personnes de 16 ans ou plus avec gains (NBEAR27), mode d'occupation (DWTENR25), état de santé (CRHLT26) et état de l'incapacité (DISABS26). Le résultat concernant la variable du mode d'occupation surprend le lecteur, car la plupart des observateurs estiment qu'il s'agit d'une variable importante pour expliquer les variations de taux d'effort. Ce résultat ne doit pas être sorti de son contexte, qui est celui d'une analyse multivariée. Les analyses transversales menées lors de cette recherche indiquent que, lorsque toutes les variables du modèle sont prises en considération (dont la variable principale source de revenu du soutien principal), il n'y a pas d'évidence que le fait d'être propriétaire ou locataire ait de l'importance sur le niveau du taux d'effort. La prise en compte de l'ensemble de ces variables fait aussi que le nombre de personnes de 16 ans ou plus avec gains dans le ménage est également non significatif. Ainsi en va-t-il des deux variables liées à la santé (incapacité et état de santé).

Finalement, il y a le groupe restant des quatre variables à résultats partagés. Trois obtiennent des résultats significatifs deux années sur trois, ce qui nous incite à les conserver pour les analyses longitudinales à venir; ce sont le sexe, le statut de déménagement (MOVST29) et le niveau d'études (HLEV2G18). Par contre, la variable indicatrice de l'immigration du répondant n'est significative qu'en 2002, et ce, à un niveau tout près du seuil de rejet (p : 0,046). Elle ne sera pas retenue pour les analyses ultérieures[39].

L'analyse multivariée des fichiers transversaux de 2002, 2003 et 2004 indique donc qu'à partir d'un ensemble de treize variables communes à ces trois années huit émergent comme significatives, c'est-à-dire qu'elles influencent le fait pour un ménage de se situer dans un rapport d'abordabilité supérieur au niveau de base (que sont les rapports d'abordabilité de moins de 30 %). Ces variables sont l'âge, le sexe, la principale source de revenu, le genre de famille, l'activité principale du soutien principal, la taille de la région de résidence, le statut de déménagement et le niveau d'études[40].

2.3 RAPPORTS DE COTES DES MODÈLES MULTIVARIÉS TRANSVERSAUX

Les rapports de cotes (r.c.) sont l'outil de travail de la régression logistique. Il s'agit de rapports entre les niveaux d'une variable indépendante et son niveau de référence (le niveau de référence est indiqué par un rapport de cotes égal à 1,0) en relation avec la variable dépendante. Le tableau 2.3 (voir plus loin) présente les rapports de cotes pour les deux équations logits. Le logit l

38. Ces observations distinctes sont produites soit par l'échantillon transversal de rafraîchissement, soit par des modifications au niveau du soutien principal du ménage.
39. Comme il est indiqué au tableau 2.1, sa relation bivariée avec la variable dépendante n'était pas significative dans le fichier transversal de 2003. Mentionnons que cette variable – indicateur de statut d'immigration – capte uniquement le fait d'être né au Canada ou pas. La durée d'immigration n'intervient pas.
40. Ces deux dernières variables, de même que celle du sexe, sont significatives deux années sur trois.

est la comparaison des ménages aux taux d'effort de 50 % et plus contre ceux de moins de 30 % (0 % à 29,9 %); le logit II est la comparaison des ménages aux taux d'effort moyennement élevés de 30 % à 49,9 % contre ceux de moins de 30 %[41].

Le tableau précédent (tableau 2.2) nous renseignait, globalement, sur les variables qui indiquent une différence significative au sein de l'échantillon. Les huit variables avec des seuils observés inférieurs à 0,05 pour la majorité des années considérées (deux années sur trois) nous disent que les ménages aux taux d'effort moyennement élevés, ou ceux avec fardeau très élevé, ou les deux, sont significativement différents du groupe de référence[42]. L'examen ultérieur des rapports de cotes observés pour ces variables nous renseigne sur les niveaux particuliers de celles-ci qui provoquent vraisemblablement le fait pour telle variable d'être significative ou non.

41. Ces équations logits indiquent que le groupe de référence de ces comparaisons sera composé des ménages sans fardeau des coûts (rapports abordables). Le choix par défaut de SUDAAN était d'utiliser la catégorie codée la plus élevée, soit les ménages aux taux d'effort de 50 % et plus. Avec ce groupe restreint de ménages comme niveau de référence, on retrouvait parfois des difficultés d'interprétation dans les données (effet principal de la variable indépendante significatif, mais aucun rapport de cotes significatif pour cette variable). Le choix de prendre le groupe de ménages le plus populeux comme groupe de référence a permis de rendre plus transparent le lien entre les effets principaux et les rapports de cotes.
42. Les ménages pour lesquels le coût d'habitation ne représente aucun fardeau.

Tableau 2.3 Fichiers transversaux 2002-2004 : rapports de cotes et seuils observés

Variables	2002				2003				2004			
	Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %		Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %		Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %	
	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)
Âge												
De 16 à 29 ans	7,40	0,0000	2,80	0,0007	5,50	0,0020	4,20	0,0000	2,40	0,0950	2,20	0,0130
De 30 à 44 ans	4,30	0,0001	2,00	0,0017	4,70	0,0001	4,20	0,0000	2,90	0,0045	2,90	0,0001
De 45 à 64 ans	4,20	0,0000	1,50	0,0410	4,40	0,0000	1,90	0,0062	3,70	0,0000	1,90	0,0020
65 ans et plus	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Sexe												
Homme	0,53	0,0124	0,60	0,0001	0,60	0,0205	0,54	0,0000				
Femme	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Principale source de revenus du ménage (3 ou 4 catégories)												
Salaires et traitements, pensions retraite et autres placements	1,00	-	1,00	-								
Salaires et traitements					0,50	0,2700	0,90	0,7400	0,80	0,7660	0,96	0,9210
Pensions retraite et autres placements					1,00	-	1,00	-	1,00	-	1,00	-
Travail autonome	12,70	0,0000	6,10	0,0000	4,30	0,0480	4,60	0,0020	13,20	0,0006	5,70	0,0006
Transferts gouvernementaux, autres revenus	11,30	0,0000	5,50	0,0000	7,60	0,0007	8,10	0,0000	10,50	0,0004	8,00	0,0000
Genre de famille (4 ou 5 catégories)												
Personne seule	5,70	0,0000	2,60	0,0000	5,80	0,0000	2,70	0,0000	5,10	0,0000	3,10	0,0000
Couple sans enfants					1,00	-	1,00	-	1,00	-	1,00	-
Couple sans enfants et autres types	1,00	-	1,00	-								
Couple avec enfants	0,76	0,5500	0,97	0,8940	0,50	0,2530	0,80	0,4890	0,45	0,1430	1,00	0,9700
Famille monoparentale	2,10	0,1120	3,00	0,0000	2,70	0,0480	2,40	0,0040	1,80	0,2290	2,80	0,0003
Autres types					0,50	0,4020	0,30	0,0060	0,65	0,6220	0,70	0,3350
Activité principale du soutien principal du ménage (4 catégories)												
Travaille	0,66	0,2380	0,70	0,0550	0,90	0,8260	0,77	0,2530	0,97	0,9140	0,82	0,4170
Va à l'école	4,10	0,0270	1,80	0,2450	10,50	0,0002	3,10	0,0123	34,00	0,0000	7,60	0,0010
Retraite, Codes réservés*	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Autres	1,20	0,5460	1,03	0,8650	1,40	0,2390	1,04	0,8690	1,80	0,0270	1,30	0,2400

* Regroupe les codes traitement provisoire, ne sait pas, refus et non applicable.

Tableau 2.3 Fichiers transversaux 2002-2004 : rapports de cotes et seuils observés (suite)

Variables	2002				2003				2004			
	Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %		Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %		Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %	
	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)
Taille de la région de résidence (5 catégories)												
Rurale	0,47	0,0040	0,62	0,0050	0,40	0,0005	0,40	0,0000	0,33	0,0000	0,42	0,0000
Urbaine : 0 à 29 999 habitants	0,43	0,0020	0,72	0,0500	0,23	0,0000	0,50	0,0004	0,41	0,0078	0,50	0,0005
Urbaine : 30 000 à 99 999 habitants	0,60	0,1140	0,82	0,3440	0,50	0,0960	0,80	0,2550	0,48	0,0250	0,74	0,1510
Urbaine : 100 000 à 499 999 habitants	0,52	0,0380	0,82	0,1900	0,60	0,1220	0,76	0,1380	0,42	0,0046	0,80	0,2350
Urbaine : 500 000 habitants et plus	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Statut de déménagement (3 catégories)												
Déménagement du ménage					2,90	0,0008	1,70	0,0082	2,00	0,0234	1,20	0,3960
Déménagement du ménage après une séparation					3,90	0,0063	1,65	0,2320	3,16	0,0160	1,80	0,1490
Ménage non déménagé, nouveau membre	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Plus haut niveau d'études du soutien principal du ménage												
Moins qu'un D.E.S.**	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Dipl. études secondaires					0,70	0,2000	0,70	0,0670	0,97	0,9230	0,65	0,0350
Postsecondaire non universitaire					0,90	0,6620	0,66	0,0390	1,33	0,3680	0,80	0,3110
Dipl. ou certificat universitaire					0,36	0,0160	0,33	0,0005	0,30	0,0150	0,57	0,0386
Codes réservés					1,56	0,2620	0,77	0,3480	1,20	0,6870	0,75	0,3410
** Diplôme d'études secondaires												
Indicateur – Personne immigrante												
Oui	1,50	0,2890	2,10	0,0002								
Non	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-
Codes réservés	1,70	0,3160	1,40	0,3600								
Indicateur – Exploitation d'entreprise à partir du ménage propriétaire												
Oui					8,20	0,0000	2,10	0,0370				
Non					1,00	-	1,00	-				
Indicateur – Autres revenus												
Rev. autr. > 0 $									0,40	0,0026	0,62	0,0220
Rev. autr. nuls									1,00	-	1,00	-

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

2.3.1 Rapports de cotes internes aux logits et aux variables

Le tableau 2.3 indique que les variables ayant un seuil observé plus élevé (plus près de zéro – voir le tableau 2.2) vont avoir tendance à présenter une situation plus claire. Ainsi, les rapports de cotes des niveaux de la variable âge montrent que les groupes d'âge sont à peu près tous significativement différents du niveau de référence, et ce, pour les deux logits. Le cas de la variable sexe est différent parce qu'il s'agit d'une variable dichotomique. Dès que le modèle indique qu'une variable de ce type est significative, le rapport de cotes entre les deux niveaux de cette variable l'est également (du moins pour l'un des logits).

Pour les autres variables, le portrait est généralement plus nuancé. Pour 2003 et 2004, le niveau de référence de la variable principale source de revenus est celui des ménages tirant la majeure partie de leurs revenus des pensions de retraite et des autres placements. Les ménages du niveau salaires et traitements ne sont pas significativement différents de ceux-ci. Par contre, les ménages tirant leur subsistance du travail autonome et ceux la tirant des transferts gouvernementaux et autres revenus le sont.

La variable genre de famille a pour niveau de référence les couples sans enfants. En général, les couples *avec* enfants ne s'en distinguent pas. Les familles monoparentales de l'équation logit II (taux d'effort de 30 % à 49,9 %) se distinguent du niveau de référence, mais pas les familles monoparentales du logit I (taux d'effort de 50 % et plus), ce qui reste difficile à expliquer[43]. La situation la plus claire est celle des personnes seules, qui se distinguent significativement du niveau de référence pour les deux logits et pour toutes les années.

La variable activité principale du soutien principal a pour niveau de référence les ménages qui sont à la retraite (et ceux qui se retrouvent sous les codes réservés), et seul le groupe des ménages dont le soutien est aux études (va à l'école) s'en distingue significativement à peu près partout. Pour les ménages qui ont d'autres activités principales, y incluant les ménages de la catégorie résiduelle autres types, on observe relativement peu de différences significatives. Cette variable est liée à la variable de la principale source de revenu; il est très probable que le niveau va à l'école aille chercher un élément explicatif non couvert par cette dernière.

La variable suivante, soit la taille de la région de résidence, illustre les limites de l'examen des pourcentages, tel que cela a été pratiqué au tableau 2.1. Les pourcentages des taux d'effort à niveau acceptable variaient de 77,3 % (régions urbaines de 100 000 à 499 999 habitants) à 86,6 % pour les régions rurales. Donc, à peine dix points de pourcentage séparent le niveau minimal du niveau maximal. Pourtant cette variable est très significative les trois années. En prenant le niveau urbain de 500 000 habitants et plus comme niveau de référence, les rapports de cotes indiquent que les contrastes significatifs sont ceux avec les deux niveaux de régions les moins densément peuplées, soit le niveau rural et le niveau urbain de 0 à 29 999 habitants. Les niveaux 30 000 à 99 999 habitants et 100 000 à 499 999 habitants n'ont pas tendance à se démarquer du niveau urbain des 500 000 et plus, du moins en ce qui concerne le logit II. Pour le logit I, on observe des rapports de cotes significatifs parmi ces deux derniers niveaux en 2002 (de 100 000 à 499 999 habitants) et en 2004 (tous les niveaux).

Les deux variables statut de déménagement et plus haut niveau d'études du soutien principal sont significatives en 2003 et 2004 seulement. Les ménages non déménagés et les ménages qui accueillent un nouveau membre sont le niveau de référence de la variable statut de

43. Comme il y a moins de ménages au logit I, il est certain que le test perd de la puissance.

déménagement. Les rapports de cotes de son logit I sont tous significatifs, ce qui n'est pas le cas du logit II, dont un seul est significatif (niveau déménagement du ménage en 2003). À l'intérieur du logit I, le niveau déménagement du ménage à la suite d'une séparation semble entraîner un rapport de cotes plus élevé que déménagement du ménage tout court.

Du côté du plus haut niveau d'études, le niveau de référence est moins qu'un diplôme d'études secondaires. Seul le rapport de cotes pour le niveau diplôme ou certificat universitaire appuie de façon soutenue le fait que la variable soit significative. Les rapports de cotes pour les autres niveaux sont inconsistants.

2.3.2 Rapports de cotes relatifs à la différenciation des ménages à fardeau moyennement élevé et à fardeau très élevé

En plus d'aider à comprendre en vertu de quels contrastes une variable est significative, l'examen des rapports de cotes est susceptible d'éclairer sur la question plus complexe et délicate de la différenciation pouvant exister parmi les ménages ayant des rapports d'abordabilité égaux ou supérieurs à 30 %. En effet, le fait qu'une variable soit significative indique qu'il y a, à tout le moins, deux groupes différents dans la population à l'étude par rapport à cette variable. Maintenant, cette étude-ci se questionne également à savoir si les ménages avec fardeau se distinguent entre eux selon qu'il s'agit d'un fardeau très élevé (logit I) ou d'un fardeau moyennement élevé (logit II). La comparaison des rapports de cotes entre les deux logits peut permettre d'éclairer cette question[44].

Cet éclairage sera tributaire de l'écart entre les rapports de cotes et de leur consistance. Pour ce qui est de l'écart, nous ne disposons pas de règle statistique. Nous allons prendre pour règle que cet écart se rapproche du double (ou de la moitié, lorsque les probabilités du niveau étudié sont moindres) en faveur du logit I, qui est le logit plus « extrême[45] », par rapport au logit II. Une indication peut-être plus décisive que nous avons affaire à deux groupes distincts de ménages avec fardeau (mais plus difficile à expliquer) serait qu'un niveau donné d'une variable présente un rapport de cotes significativement inférieur pour un logit, et significativement supérieur pour l'autre logit, bref que ceux-ci évoluent en direction opposée.

La variable âge présente plusieurs cas de rapports de cotes du logit I qui doublent ceux du logit II, du moins pour l'année 2002. Pour les années 2003 et 2004, on ne l'observe que pour le groupe d'âge des 45 à 64 ans.

La variable sexe est neutre à cet égard.

La variable principale source de revenus répète les résultats de la variable âge pour l'année 2002, c'est-à-dire que les rapports de cotes du logit I sont le double de ceux du logit II. Cet écart s'estompe en 2003, reprend en 2004, mais seulement pour le niveau de travail autonome (rapports de cotes logit I = 13,2; celui du logit II = 5,7).

La variable genre de famille donne une indication consistante pour la catégorie de famille « personne seule », dont les rapports de cotes du logit I sont à peu près le double de ceux du logit II pour les trois années. Dans les autres catégories, il n'y a rien de significatif ni de consistant.

La même situation se répète pour la variable activité principale du soutien principal. Lorsque l'activité du soutien principal est de fréquenter l'école, les rapports de cotes du logit I sont au moins

44. Il s'agit bien d'un éclairage, car l'idéal aurait été que l'on dispose de n suffisants pour former une équation logit comparant directement les ménages à taux d'effort de 30 % à 49,9 % à ceux de 50 % et plus.
45. Le logit I compare les taux d'effort de 50 % et plus au niveau de référence (taux d'effort de moins de 30 %).

le double des rapports de cote du logit II. Pour les autres catégories, à peu près aucun rapport de cotes n'est significatif.

La variable taille de la région de résidence ne présente rien qui se distingue de façon marquée. Le logit I a tendance à offrir des rapports de cotes inférieurs à ceux du logit II, mais seul le niveau « Urbain : 0 à 29 999 habitants » présente un écart d'environ la moitié, et ce, pour les années 2002 et 2003 seulement. L'autre niveau significativement différent du niveau de référence est le niveau rural, mais les écarts de rapports de cotes entre logits y sont réduits.

La variable statut de déménagement est peut-être la variable qui présente le cas le plus convaincant, jusqu'à maintenant, d'une différenciation plus poussée parmi les ménages avec fardeau des coûts. En effet, les quatre rapports de cotes rattachés au logit I sont significatifs, alors qu'un seul (sur quatre) l'est pour le logit II, soit celui de l'année 2003 de la catégorie déménagement du ménage. Sa valeur est de 1,7, alors que la valeur de son vis-à-vis au logit I est de 2,9. Pas le double, mais nettement supérieur quand même.

Quant à la variable du plus haut niveau d'études, il n'y a qu'un seul niveau qui est significatif et consistant par rapport au niveau de référence, soit le niveau diplôme et certificat universitaire. Les rapports de cotes pour 2003 sont pratiquement à égalité, mais ceux pour 2004 montrent un écart intéressant. Pour ceux qui détiennent un diplôme ou un certificat universitaire, le rapport de cotes pour le logit I, c'est-à-dire l'événement de connaître un taux d'effort de 50 % et plus contre moins de 30 %, indique des probabilités inférieures d'environ la moitié (rapport de cotes = 0,30) à son équivalent pour le logit II (rapport de cotes = 0,57)[46].

Nous ne retrouvons pas dans les données de cas où les rapports de cotes pour un niveau donné d'une variable vont en direction opposée dans l'un et l'autre logit. Par contre, on retrouve quelques cas où le rapport de cotes pour le logit I, qui représente la population plus « extrême », est non significatif, alors que son vis-à-vis pour le logit II l'est. Cela tend à invalider l'hypothèse d'une différenciation soutenue parmi les ménages avec fardeau des coûts. Ces cas sont :

Tableau 2.4 Variables, niveaux et années pour lesquels le rapport de cotes du logit I est non significatif, contrairement à celui du logit II

Variable	Niveau	Année
Âge	16 à 29 ans	2004
Genre de famille	Famille monoparentale	2002
Genre de famille	Famille monoparentale	2004
Plus haut niveau d'études	Diplôme d'études secondaires	2004
Plus haut niveau d'études	Postsecondaire non universitaire	2003

Ces cas sont quand même peu nombreux. Le cas le plus préoccupant est peut-être celui de la variable genre de famille, d'abord parce qu'il se produit pour deux années, et ensuite parce qu'il avait été mentionné plus tôt que les ménages de famille monoparentale et ceux de personne seule sont les deux niveaux de cette variable qui se distinguent le plus du niveau de référence. On

46. Un tableau d'ensemble de ces cas d'écart des rapports de cotes entre les logits I et II figure à l'annexe 7.

constate que cette affirmation a besoin d'être atténuée en ce qui concerne les familles monoparentales[47].

CONCLUSION

L'objectif poursuivi dans ce chapitre était de développer un modèle général initial à l'aide des fichiers transversaux de l'EDTR, et ce, avant d'entamer les analyses longitudinales. À partir d'un ensemble de treize variables indépendantes, nous considérons que huit contribuent significativement à éclairer les différences entre les ménages avec taux d'effort acceptable et ceux enregistrant un taux d'effort pouvant être qualifié de fardeau.

Cinq variables deviennent donc moins prioritaires, puisque le modèle atténue leur lien avec la variable dépendante. Ce sont le nombre de personnes de 16 ans ou plus avec gains (nbear27), le mode d'occupation (dwtenr25), l'état de santé (crhlt26), l'état de l'incapacité (disabs26) et le statut d'immigration (immst15). De ce groupe de variables, seuls l'état de santé et le statut d'immigration seront tenus à l'écart dans les modèles longitudinaux.

Les variables à considérer en priorité sont l'âge, le sexe, la principale source de revenus (mjsif27), le genre de famille économique (fmcomp27), l'activité principale du répondant principal (mjacg27), la taille de la région de résidence (urbszg25), le statut de déménagement (movst29) et le niveau d'études (hlev2g18).

Un niveau de référence est déterminé pour chacune de ces variables et ensuite on observe les écarts des autres niveaux par rapport à celui-ci (à l'aide des rapports de cotes). Pour la variable âge, le niveau de référence est le groupe d'âge des 65 ans et plus. Les autres groupes d'âge enregistrent tous des probabilités supérieures de connaître des taux d'effort avec fardeau, et ce, pour les trois années de suivi.

La variable sexe est significative deux années sur trois. Elle indique que les hommes obtiennent des probabilités inférieures de connaître des taux d'effort avec fardeau.

La variable principale source de revenu du ménage a pour niveau de référence les pensions de retraite et les autres placements. Les ménages ayant les salaires et traitements comme principale source ne s'en distinguent pas. Toutefois, les ménages qui tirent leur subsistance du travail autonome ou des transferts gouvernementaux obtiennent des probabilités nettement supérieures de connaître des taux d'effort avec fardeau.

La variable genre de famille a pour niveau de référence les couples sans enfants. Ce sont les ménages de personnes seules qui s'en distinguent le plus nettement. Ils présentent des probabilités plus fortes de connaître des rapports d'abordabilité avec fardeau. Les familles monoparentales suivent dans ce classement. Leurs probabilités de connaître un fardeau modéré (30 % à 49,9 %) par rapport aux couples sans enfants sont comparables à celles des personnes seules, mais pas leurs probabilités de connaître un fardeau élevé (50 % et plus), qui sont non significatives deux années sur trois.

La situation de la variable activité principale du soutien principal est fort simple. Les ménages à la retraite et ceux avec codes réservés forment le niveau de référence, et les seuls ménages qui s'en distinguent significativement sont les ménages dont l'activité principale du soutien principal est de

47. Un n fragile pourrait contribuer à ce résultat pour ce genre de ménages, puisqu'il n'y avait que 35 ménages monoparentaux avec fardeau élevé en 2003 (d'après le tableau 2.1).

fréquenter l'école. Comme on l'imagine toutefois, ces ménages dirigés par une personne aux études ne forment qu'un groupe restreint d'à peine 2 % de l'échantillon (112 ménages en 2003).

Ces fichiers transversaux nous enseignent donc que, parmi un ensemble de treize variables avec potentiel, les ménages avec fardeau des coûts d'habitation se distinguent des ménages sans fardeau des coûts par rapport à huit de ces variables. Ce chapitre s'est aussi intéressé à la question de savoir si les ménages avec fardeau des coûts se distinguent entre eux selon que le fardeau est modéré (30 % à 49,9 %) ou élevé (50 % et plus). C'est là une question plus délicate que nous n'avons que commencé à aborder.

Les indications fournies par les fichiers transversaux sont, globalement, que les rapports de cotes entre les logits I et II vont dans le même sens. Il n'y a pas de trajectoires opposées. Tout au plus relève-t-on un nombre limité de cas où le rapport de cotes pour le logit I est non significatif, alors que le rapport de cotes pour le logit II l'est. On observe toutefois un nombre plus élevé de cas où le rapport de cotes pour le logit I est plus accentué (double, ou moitié moins) que celui pour le logit II.

Cela, on l'observe pour certains niveaux de la variable âge, tel celui du groupe d'âge des 45 à 64 ans, pour le niveau personne seule de la variable genre de famille (les trois années), pour l'activité principale d'aller à l'école et, finalement, pour la variable statut de déménagement, pour laquelle le niveau déménagement du ménage à la suite d'une séparation n'est vraiment significatif qu'au logit I.

Dans les analyses poursuivies dans ce chapitre, nous avons traité les fichiers des trois années sous observation comme s'ils étaient indépendants. Il s'agit en fait d'une enquête longitudinale et les fichiers des trois années sont interreliés. Il y a environ 5 800 enregistrements par année. Or le fichier longitudinal à proprement parler, que nous aborderons au prochain chapitre, va compter 4 832 enregistrements. La variabilité des résultats d'une année à l'autre dans les tableaux 2.2 et 2.3 est donc attribuable en bonne partie aux quelque mille enregistrements qui s'ajoutent ponctuellement pour une année donnée.

CHAPITRE 3 : PREMIER REGARD SUR LES DONNÉES LONGITUDINALES

3.1 LE PORTRAIT DES CHEMINEMENTS EN MATIÈRE DE TAUX D'EFFORT

Après définition d'un certain nombre de critères d'inclusion[48], SLIDRET (le logiciel d'extraction spécifique à l'EDTR) a bâti un fichier de 4 832 enregistrements longitudinaux. Ce fichier compte moins d'enregistrements que les fichiers transversaux parce qu'il ne retient que les ménages qui sont répondants à l'enquête durant l'ensemble de la période (2002 à 2004)[49].

Le chapitre 3 aborde la question de la stabilité dans le temps des taux d'effort des ménages (la variable dépendante). En premier lieu, nous livrerons un tableau strictement descriptif des cheminements en question. Le cheminement des ménages québécois du fichier longitudinal sera synthétisé par un vecteur de trois chiffres accolés, chiffres compris entre 1 et 3. Chacun de ces chiffres recouvre deux significations, soit l'année et le taux d'effort.

Le tableau suivant illustre la signification de ces vecteurs.

Tableau 3.1 Représentation < nombre et indice > des vecteurs de taux d'effort [n_1,n_2,n_3]

Nombre	Année
En position n_1	2002
En position n_2	2003
En position n_3	2004
Indice	**Taux d'effort**
Indice 1	0 % à 29,9 %
Indice 2	30 % à 49,9 %
Indice 3	50 % et plus

48. Être soutien principal du ménage et résider au Québec, revenu du ménage ≥ 6 500 $, coûts d'habitation < 100 000 $. Ces spécifications s'appliquent à l'année de clôture (2004). En plus, rapport d'abordabilité (taux effort) plus grand ou égal à 0 pour chacune des années.
49. L'attrition de l'échantillon est prise en compte par la pondération. Il est à noter que l'identité du soutien économique principal du ménage peut changer en cours de période. Conséquemment, des variables telles que l'âge et le sexe de celui-ci ne sont pas fixes dans le temps.

Le tableau qui suit contient les proportions estimées des cheminements de l'ensemble des ménages au cours des trois années sélectionnées.

Tableau 3.2 Vecteurs des cheminements des ménages, Québec, 2002-2004, données pondérées

			Cumulatif	
Vecteur	**Nombre**	**%**	**Nombre**	**%**
111	3 597,9	74,5	*3 597,9*	*74,5*
112	147,0	3,0	*3 744,9*	*77,5*
113	41,1	0,9	*3 786,0*	*78,4*
121	86,4	1,8	*3 872,4*	*80,1*
122	83,0	1,7	*3 955,4*	*81,9*
123	23,9	0,5	*3 979,3*	*82,4*
131	28,9	0,6	*4 008,2*	*82,9*
132	17,8	0,4	*4 025,9*	*83,3*
133	17,6	0,4	*4 043,5*	*83,7*
211	175,9	3,6	*4 219,4*	*87,3*
212	27,7	0,6	*4 247,1*	*87,9*
213	14,7	0,3	*4 261,8*	*88,2*
221	58,9	1,2	*4 320,7*	*89,4*
222	191,1	4,0	*4 511,9*	*93,4*
223	23,9	0,5	*4 535,7*	*93,9*
231	6,7	0,1	*4 542,4*	*94,0*
232	24,0	0,5	*4 566,4*	*94,5*
233	18,6	0,4	*4 585,1*	*94,9*
311	54,4	1,1	*4 639,5*	*96,0*
312	6,2	0,1	*4 645,7*	*96,1*
313	7,2	0,1	*4 652,9*	*96,3*
321	8,4	0,2	*4 661,4*	*96,5*
322	40,3	0,8	*4 701,6*	*97,3*
323	11,3	0,2	*4 712,9*	*97,5*
331	9,6	0,2	*4 722,5*	*97,7*
332	24,2	0,5	*4 746,6*	*98,2*
333	85,5	1,8	*4 832,1*	*100,0*

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

Il existe une possibilité théorique de 27 cheminements, et on retrouve des ménages dans chacun de ces cheminements. Ce qui ressort toutefois, c'est que, sur cette période de trois ans, les trois quarts des ménages québécois sont demeurés avec des taux d'effort acceptables. Ce sont les ménages du vecteur [1 1 1].

Le quart des ménages québécois ont donc connu au moins une période d'une année avec un taux d'effort de 30 % ou plus, qu'il s'agisse d'un fardeau moyennement élevé (30 % à 49,9 %) ou d'un fardeau très élevé (50 % et plus). Parmi les 26 cheminements autres que le [1 1 1], trois recueillent un effectif égal ou supérieur à 3 %. Ce sont les cheminements [1 1 2] (3,0 %), [2 1 1] (3,6 %) et [2 2 2] (4,0 %). Cinq cheminements se situent entre 1 % et 2 %, et tous les autres (18) sont inférieurs à 1 %.

Pour obtenir un deuxième regard complémentaire sur ces cheminements, on peut consulter la matrice de corrélation que l'on obtient avec les taux d'effort annuels[50]. Il est à noter que les trois cheminements parfaitement corrélés [1 1 1; 2 2 2; 3 3 3] totalisent un peu plus de 80 % des ménages. Pourtant, les coefficients de corrélation pondérés obtenus sont les suivants.

Tableau 3.3 Coefficients de corrélation des taux d'effort pour la période 2002-2004 (et seuils observés), Québec, données pondérées

Coefficients de corrélation r, N=4832 Prob > |r| sous l'hypothèse nulle : R=0	Taux_2002	Taux_2003	Taux_2004
Coefficients r de taux_2002 avec... (Niveau de taux d'effort en 2002) Seuils observés	1,00000	0,59078 < 0,0001	0,52722 < 0,0001
Coefficients r de taux_2003 avec... (Niveau de taux d'effort en 2003) Seuils observés	0,59078 < 0,0001	1,00000	0,65221 < 0,0001
Coefficients r de taux_2004 avec... (Niveau de taux d'effort en 2004) Seuils observés	0,52722 < 0,0001	0,65221 < 0,0001	1,00000

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

Avec des données de cette nature, on constate donc que des coefficients supérieurs à 0,50, tels qu'ils se trouvent, dénotent une assez forte corrélation. On remarque également que les corrélations sur deux années successives (2002 et 2003, 2003 et 2004) sont légèrement supérieures à la corrélation entre 2002 et 2004, ce qui indique une légère diminution dans le temps de la corrélation des taux d'effort.

Jusqu'à maintenant, ce chapitre s'est penché sur la synthèse des cheminements par rapport aux variables dépendantes (taux d'effort ou rapports d'abordabilité). Il faut maintenant examiner quelles variables influencent significativement les taux d'effort des ménages sur une période de trois ans.

3.2 LE MODÈLE LONGITUDINAL AVEC FICHIER DE MESURES RÉPÉTÉES

Pour répondre à cette question, la méthode usuelle consiste d'abord à restructurer le fichier de 4 832 observations, contenant toutes les variables pour chacune des années, en un fichier de 14 496 (4 832 x 3) observations, avec une observation par ménage par année. Chaque ménage génère donc trois observations. Une nouvelle variable indépendante est créée pour indiquer l'année : la variable temps (1 = 2002; 2 = 2003; 3 = 2004).

La méthode d'analyse la plus adéquate consiste à requérir de SUDAAN l'exécution d'une analyse de régression logistique multinomiale sur des observations corrélées longitudinalement, donc en

50. Corrélations de Pearson avec SAS. Ce logiciel n'autorise pas l'utilisation de la pondération avec le coefficient de corrélation de Spearman, qui aurait peut-être été plus indiqué.

spécifiant une structure de corrélation dite échangeable[51]. Ce type d'analyses a été mené. Malheureusement, malgré la longueur des temps d'exécution, SUDAAN n'a pu faire converger les modèles testés.

L'autre solution qui s'offrait à nous était de considérer les observations répétées (2002, 2003, 2004) comme étant indépendantes longitudinalement. C'est ce qui a été suggéré par un méthodologiste de Statistique Canada, puisque « l'expérience montre que le choix d'une matrice de corrélation autre que l'identité (c'est-à-dire l'indépendance) n'améliore pas beaucoup l'analyse[52] ».

Avec cet arrangement pragmatique, qui faisait fi de la corrélation longitudinale, la procédure de SUDAAN a convergé.

Au point d'entrée, le modèle à mesures répétées a repris la grande majorité des variables incluses dans les modèles transversaux, c'est-à-dire toutes les variables qui y furent significatives, et trois des quatre variables qui y furent non significatives. Celles-ci sont le nombre de personnes de 16 ans ou plus avec gains, le mode d'occupation et l'état de l'incapacité. La variable non significative laissée de côté a été la variable état de santé, puisqu'en matière de santé l'étude comptait déjà état de l'incapacité. Comme il a été mentionné au chapitre 2, la variable indicateur d'immigration, significative qu'une année sur trois, n'a pas été ramenée non plus[53].

Cette méthode légèrement plus parcimonieuse s'explique par la nécessité d'« allouer de l'espace » à la variable temps, qu'un modèle de mesures répétées amène à intégrer, de même qu'aux interactions entre cette variable et chacune des variables indépendantes significatives. En effet, ces interactions devront être testées pour s'assurer que l'effet de ces variables indépendantes puisse être considéré comme stable dans le temps. De cette manière, on peut en arriver à un modèle mieux spécifié.

Le tableau 3.4 présente les résultats initiaux du modèle de régression logistique avec mesures répétées. Ce modèle donne toutes les variables utilisées et reprises des modèles transversaux comme significatives, sauf les variables incapacité et mode d'occupation. La variable nombre de personnes avec des gains est faiblement significative ($p = 0,043$). Finalement, la variable spécifique temps est non significative[54].

51. Syntaxe Sudaan : option R = exchangeable. L'hypothèse émise est que les corrélations entre deux observations à deux points dans le temps sont égales.
52. M. Abdelnasser Saidi, courriel du 8 mars 2007 adressé à Danielle Forest.
53. Les variables testées de façon multivariée d'une année seulement ont également été laissées de côté, soit les variables autres revenus et ménage exploite une entreprise.
54. On peut interpréter ce résultat en soumettant que l'année considérée n'affecte pas la probabilité qu'un ménage ait un taux d'effort égal ou supérieur à 30 % (un taux d'effort avec fardeau des coûts).

Tableau 3.4 Modèle – Mesures répétées, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 (non significatif : seuil observé ≥ 0,05)

Variables	Seuils observés
Âge	0,0000
Sexe	0,0184
Princ. source revenu	0,0000
Genre de famille	0,0000
Activité principale	0,0000
Taille région résidence	0,0000
Déménagement	0,0000
Niveau d'études	0,0006
Incapacité	Non significatif
Mode occupation	Non significatif
Nbre pers. avec gains	0,0427
Temps	Non significatif

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal.
Compilation : Société d'habitation du Québec.

La non-significativité de cette dernière variable (temps) ne dispense pas de tester les interactions entre celle-ci et les variables indépendantes significatives. Ces interactions ont été ajoutées une à la fois. Autant de modèles qu'il y a de variables significatives (9) ont été soumis à SUDAAN. Le tableau 3.5 rapporte les résultats des interactions.

Tableau 3.5 Modèle – Mesures répétées, seuils observés des interactions temps*variables indépendantes, fichier longitudinal 2002-2004

Interactions	Seuils observés
Temps*Âge	0,9470
Temps*Sexe	0,6050
Temps*Princ. source revenu	0,2140
Temps*Genre de famille	0,6570
Temps*Activité principale	0,2270
Temps*Taille région résidence	0,8850
Temps*Déménagement	0,9010
Temps*Niveau d'études	0,1950
Temps*Nbre pers. avec gains	0,4280

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec .

Ce tableau indique qu'aucune interaction n'est significative[55]. L'effet des variables indépendantes peut donc être considéré stable dans le temps. Les résultats de ce modèle peuvent également être comparés sommairement aux résultats transversaux, en gardant présent toutefois que le modèle avec mesures répétées est avantagé par sa taille échantillonnale.

55. Le seuil observé de certaines interactions, notamment l'interaction temps*genre de famille, et, dans une moindre mesure, celle du temps*principale source de revenu, peut être considéré comme moins précis que ce qui est indiqué au tableau, étant donné le nombre de poids de réplication pour lesquels SUDAAN a émis des avertissements (« The data may have singularities for the model you are trying to fit [...] »).

Le fait pour une variable d'être significative doit se traduire, normalement, par au moins un rapport de cotes dont le seuil observé est sous le seuil de signification ($\alpha < 0,05$). Le tableau 3.6 rapporte les rapports de cotes obtenus pour le modèle longitudinal.

Tableau 3.6 : Rapports de cotes, longitudinal 2002-2004 – Mesures répétées

Variables	Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %	
	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)
Âge				
De 15 à 29 ans	3,99	0,0000	2,57	0,0000
De 30 à 44 ans	3,75	0,0000	2,88	0,0000
De 45 à 64 ans	4,59	0,0000	2,10	0,0000
65 ans et plus	1,00	-	1,00	-
Sexe				
Homme	0,74	0,1100	0,73	0,0100
Femme	1,00	-	1,00	-
Principale source de revenus du ménage (4 catégories)				
Salaires et traitements	0,76	0,5900	0,93	0,7900
Travail autonome	9,51	0,0000	6,06	0,0000
Transferts gouvernementaux, autres revenus	6,93	0,0000	5,77	0,0000
Pensions retraite et autres placements	1,00	-	1,00	-
Temps				
An 2002		*n.s.*		*n.s.*
An 2003		*n.s.*		*n.s.*
An 2004	1,00	-	1,00	-
Nombre de membres de la famille économique âgés de 16 ans et plus avec gains (3 catégories)				
Aucun membre	3,18	0,0000	1,51	0,1200
Un membre	1,58	0,1300	1,12	0,5800
Deux membres ou plus	1,00	-	1,00	-
Mode d'occupation				
Propriétaire		*n.s.*		*n.s.*
Locataire	1,00	-	1,00	-
Genre de familles (5 catégories)				
Personne seule	4,08	0,0000	2,70	0,0000
Couple sans enfants	1,00	-	1,00	-
Couple avec enfants	0,85	0,6900	1,01	0,9700
Famille monoparentale	2,01	0,0800	2,97	0,0000
Autres types	0,48	0,1300	0,74	0,2400
Activité principale du soutien principal du ménage (4 catégories)				
Travaille	1,45	0,2100	0,92	0,6200
Va à l'école	8,38	0,0000	2,95	0,0000
Retraite, codes réservés*	1,00	-	1,00	-
Autres	1,48	0,0400	0,94	0,6400
* Regroupe les codes traitement provisoire, ne sait pas, refus et non applicable.				
Taille de la région de résidence (5 catégories)				
Rurale	0,33	0,0000	0,48	0,0000
Urbaine : 0 à 29 999 habitants	0,29	0,0000	0,58	0,0000
Urbaine.: 30 000 à 99 999 habitants	0,48	0,0000	0,70	0,0100
Urbaine.: 100 000 à 499 999 habitants	0,59	0,0200	0,81	0,1300
Urbaine.: 500 000 habitants et plus	1,00	-	1,00	-

Tableau 3.6 : Rapports de cotes, longitudinal 2002-2004 – Mesures répétées (suite)

Variables	Taux d'effort de 50 % et plus contre 0 % - 29,9 %		Taux d'effort de 30 % - 49,9 % contre 0 % - 29,9 %	
	Rapport de cotes	Seuils observés (Beta)	Rapport de cotes	Seuils observés (Beta)
Indicateur – État de l'incapacité pour l'année de référence				
Incapacité		*n.s.*		*n.s.*
Non-incapacité	1,00	-	1,00	-
Codes réservés		*n.s.*		*n.s.*
Statut de déménagement (3 catégories)				
Déménagement du ménage	2,24	0,0000	1,48	0,0000
Déménagement du ménage après une séparation	3,47	0,0000	1,87	0,0000
Ménage non déménagé, nouveau membre	1,00	-	1,00	-
Plus haut niveau d'études du soutien principal du ménage				
Moins qu'un D.E.S.**	1,00	-	1,00	-
Dipl. études secondaires	0,97	0,8600	0,68	0,0100
Postsecondaire non universitaire	0,78	0,3400	0,71	0,0300
Dipl. ou certificat universitaire	0,44	0,0200	0,40	0,0000
Codes réservés	1,73	0,1000	0,90	0,6700

** Diplôme d'études secondaires

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

La variable temps et ses interactions avec les autres variables indépendantes ayant été non significatives, le modèle longitudinal avec fichier de mesures répétées peut être considéré comme une superposition des modèles transversaux, en ce qui a trait aux observations communes aux trois années. Pour rendre la comparaison avec les modèles transversaux plus simple (voir le tableau 2.3), nous allons nous en tenir surtout à l'année 2003, qui est l'année du milieu de période. La similitude des rapports de cotes et des seuils observés est très grande. Cela est manifeste pour la variable âge, significative à 0,0000 (<0,0001) dans les deux modèles, et dont tous les rapports de cotes sont également significatifs.

Dans le modèle longitudinal avec fichier de mesures répétées, il y a atténuation de la variable sexe par rapport à son résultat de 2003[56], et cela se traduit par un rapport de cotes non significatif pour le logit I[57], mais significatif pour le logit II (dans le modèle transversal, les rapports de cotes pour les deux logits étaient significatifs).

Pour la variable principale source de revenus, il y a une très forte similitude. L'analyse avec fichier de mesures répétées indique que les rapports de cotes des niveaux travail autonome et transferts gouvernementaux sont ceux qui s'écartent le plus du niveau de référence (pensions de retraite et autres placements). Les seuils observés se situent plus loin dans la zone de rejet que ce qu'ils montraient dans le modèle transversal.

La variable nombre de membres de la famille économique de 16 ans ou plus avec des gains est faiblement significative (seuil observé de 0,043). Le seul rapport de cotes qui est significatif est celui pour le niveau aucun membre avec des gains, et ce, pour le logit I seulement. Cette variable n'était pas significative dans le modèle transversal de 2003. Sa significativité ici pourrait être

56. Seuil observé de la variable sexe égal à 0,0186 sous le modèle longitudinal avec fichier de mesures répétées, alors que son seuil dans le modèle transversal 2003 était 0,0001. Cette atténuation peut s'expliquer par l'année 2004, pour laquelle la variable sexe était non significative.
57. Rappel : logit I = comparaisons des ménages avec taux d'effort de 50 % et plus contre ceux avec taux de 0 % à 29,9 %; logit II : comparaisons des ménages avec taux d'effort entre 30 % et 49,9 % contre ceux de 0 % à 29,9 %.

attribuable au fait que la structure de corrélation des ménages n'est pas prise en compte dans le modèle à mesures répétées (sans oublier, bien sûr, la taille échantillonnale).

La variable mode d'occupation (logement possédé ou loué) est non significative, une répétition du résultat du modèle transversal de 2003.

La variable genre de familles donne aussi des résultats similaires au modèle transversal. Le niveau personnes seules reste celui qui se détache le plus nettement du niveau de référence (les couples sans enfants). Le niveau famille monoparentale n'est pas significatif pour le logit I, mais il l'est pour le logit II, ce qui reproduit un des résultats transversaux obtenus (celui de l'année 2004).

La variable activité principale du soutien principal du ménage donne des résultats similaires. Seul le niveau « va à l'école » se détache significativement du niveau de référence (retraite, codes réservés), avec des rapports de cotes très similaires à ceux du modèle transversal (de 2003).

Le modèle longitudinal appliqué à la variable taille de la région de résidence donne des résultats un peu différents du modèle transversal. Tous les niveaux obtiennent des résultats significatifs par rapport au niveau de référence pour le logit I, et un seul n'est pas significatif pour le logit II (soit le niveau urbain : 100 000 à 499 999 habitants). Dans les modèles transversaux, seuls les niveaux région rurale et région urbaine de moins de 30 000 habitants étaient significatifs pour toutes les années et tous les logits.

Dans le modèle longitudinal, la variable santé retenue (soit « indicateur : état de l'incapacité ») n'est pas significative, pas plus qu'elle ne l'était dans les fichiers transversaux.

Sous le modèle longitudinal avec fichier de mesures répétées, la variable statut de déménagement est significative à un seuil observé de 0,0000 (ou <0,0001), un niveau plus bas (significatif) que celui des modèles transversaux. Conséquemment, on observe que les rapports de cotes des niveaux déménagement du ménage et déménagement du ménage à la suite d'une séparation deviennent tous significatifs (à un seuil observé de 0,0000 également), ce qui n'était pas vraiment le cas dans les fichiers transversaux de 2003 et de 2004, spécialement pour le logit II (notons, de plus, que cette variable était non significative en 2002).

Finalement, il ne reste que la variable du plus haut niveau d'études du soutien principal, qui, sous le modèle transversal, était significative en 2003 et en 2004. Elle est significative à 0,0006 dans le modèle longitudinal, ce qui est plus bas (plus significatif) que son meilleur seuil observé transversal (celui de 2003). Du côté des rapports de cotes, le niveau diplôme ou le certificat universitaire est le seul qui se démarque significativement du niveau de référence dans le logit I. Cela était aussi le cas en 2003. Pour le logit II cependant, trois niveaux se démarquent significativement, soit un de plus qu'en 2003. On dénombre, en effet, diplôme d'études secondaires, postsecondaire non universitaire et, bien sûr, diplôme ou certificat universitaire.

CONCLUSION

Le modèle de mesures répétées bâtit un fichier qui est, en l'absence d'interactions significatives avec la variable temps, un amalgame des trois fichiers transversaux. On ne devrait donc pas s'étonner que la comparaison de ses résultats avec le fichier transversal de 2003 donne, *grosso modo*, des résultats convergents. Chacune des variables significatives en 2003 révèle également qu'elle influence significativement les taux d'effort sur une période de trois ans. Le modèle longitudinal procure des résultats plus généraux que chacun des fichiers transversaux individuellement.

Par la comparaison avec l'année 2003, on note tout de même de légères variantes. Une seule variable voit son importance être atténuée quelque peu : c'est la variable sexe. Le seuil observé longitudinal de 0,018 reflète probablement le fait que cette variable a été non significative en 2004.

Quatre variables nous paraissent renforcées dans le modèle longitudinal. Le renforcement est léger pour la variable principale source de revenus, puisqu'elle était déjà hautement significative pour les fichiers transversaux. On observe toutefois une accentuation des rapports de cotes pour les niveaux transferts gouvernementaux (et autres revenus) et travail autonome par rapport au niveau de référence.

Une autre variable pour laquelle on observe un renforcement relatif aux rapports de cotes est la variable taille de la région de résidence. À l'exception du niveau urbain : 100 000 à 499 999 habitants pour le logit II, les rapports de cotes pour tous les autres niveaux (et pour les deux logits) deviennent significativement différents du niveau de référence.

La troisième variable renforcée est la variable statut de déménagement. Ici, le renforcement se reflète sur le seuil observé de la variable. Avec un seuil observé de p = 0,0000, cette variable obtient un seuil plus bas que son meilleur seuil transversal (soit en 2003, seuil de 0,0009). Ce renforcement se reflète aussi par des rapports de cotes tous significatifs pour les niveaux de déménagement (déménagement tout court et déménagement à la suite d'une séparation).

Enfin, comme nous l'avons mentionné précédemment, la variable du plus haut niveau d'études est également renforcée dans le modèle longitudinal avec fichier de mesures répétées. Ce renforcement se reflète aussi sur son seuil observé, de même que sur ses rapports de cotes, particulièrement pour le logit II.

Le fichier de mesures répétées laisse les autres variables relativement en même position qu'en 2003, qu'elles soient significatives (âge, genre de familles, activité principale du soutien principal) ou non significatives (mode d'occupation, état de l'incapacité). Il y a une petite exception du côté de la variable nombre de personnes de 16 ans ou plus avec des gains, qui, de non significative dans les modèles transversaux, devient faiblement significative au seuil observé de 0,043.

Il est possible que l'augmentation de la taille de l'échantillon que provoque la constitution de ce fichier longitudinal contribue à ces renforcements observés auprès de certaines variables et des rapports de cotes entre niveaux. Puisque les tailles d'échantillons varient, il faut rester prudent dans l'interprétation des comparaisons de seuils entre les modèles longitudinal et transversaux.

Dans la première section de ce chapitre, nous avons fait le point sur les parcours longitudinaux en matière de taux d'effort observés chez les ménages au cours de cette période de trois ans. Les trois quarts des ménages sont stables, avec un taux d'effort sans fardeau des coûts (taux inférieurs à 30 %). Leur vecteur est de [1 1 1]. Au deuxième rang se retrouvent les ménages du vecteur [2 2 2], qui comptent pour 4 % de l'échantillon. Le troisième groupe en importance comprend les ménages du vecteur [2 1 1], qui totalisent 3,6 % des ménages. Les ménages restants se répartissent sur 24 autres parcours. Un des objectifs du prochain chapitre (qui traitera d'autres modèles longitudinaux) sera de mieux comprendre la structure de ces parcours.

CHAPITRE 4 : MODÈLES LONGITUDINAUX AVEC DIMENSION TEMPS INHÉRENTE

INTRODUCTION

Le chapitre précédent a tracé les cheminements longitudinaux des ménages en matière de rapport d'abordabilité et a fourni une première analyse longitudinale dans laquelle la dimension temps était prise en charge en tant que variable indépendante du modèle. À ce titre, il avait fallu éclater les enregistrements selon le nombre d'années d'observation (3 ans).

Dans ce chapitre-ci, la dimension temps va être inhérente aux modèles. Elle sera incorporée dans ce que l'on cherche à expliquer, c'est-à-dire la variable dépendante. Le nombre d'enregistrements revient donc au nombre originel de 4 832. Deux modèles longitudinaux avec dimension temps inhérente ont été testés. Dans le premier de ces modèles, l'intervention discrétionnaire du chercheur est davantage poussée, en ce sens que plus d'éléments du modèle sont contrôlés par lui. Nous allons l'appeler le modèle de taux d'effort élevé persistant (où le chercheur définit ce qu'il considère comme persistant).

Le deuxième modèle longitudinal avec dimension temps inhérente s'appellera le modèle longitudinal par trajectoire. Ce type de modèles découle d'une famille de modèles appelée, en statistiques, modèles de mélange fini. Le chercheur choisit quelques grands paramètres et laisse ensuite le modèle établir le partage des observations selon la méthodologie qui lui est propre.

4.1 MODÈLE DE TAUX D'EFFORT ÉLEVÉ PERSISTANT

Une forme d'analyse longitudinale différente de celle qui a été abordée au chapitre précédent consiste, pour le chercheur, à qualifier le taux d'effort élevé qu'il entend étudier. Cette qualification tient compte du taux d'effort et de la durée de séjour entre les niveaux de taux d'effort.

L'analyse qui a été menée de cette manière consiste d'abord à définir ce que l'on entend par taux d'effort élevé. Il a été décidé que cette définition serait la plus large possible. C'est donc dire que tous les rapports d'abordabilité égaux ou supérieurs à 30 % seront considérés élevés. La distinction entre fardeau des coûts moyennement élevé (rapports de 30 % à 49,9 %) et fardeau très élevé (50 % et plus) est mise à l'écart.

Ensuite, il faut poser un critère sur la durée. Sommes-nous satisfaits de distinguer un ménage dès qu'il connaît une année de taux d'effort élevé? Étant donné que nous avons adopté une définition large au niveau des rapports d'abordabilité, nous croyons que nous pouvions nous montrer plus exigeants au niveau de la durée. Nous voulons caractériser les ménages qui connaissent des taux d'effort élevés persistants. Un an ne répond pas au critère de persistance. Trop de circonstances particulières peuvent expliquer une mauvaise année pour un ménage. Nous allons poser qu'un taux d'effort élevé persistant consiste à connaître un rapport d'abordabilité égal ou supérieur à 30 % pendant deux ans ou trois ans (sur les trois années de la période d'observation). Donc les ménages qui seront caractérisés sont les ménages avec rapports d'abordabilité égal ou supérieur à 30 % deux ans sur trois ou durant les trois années. Quatorze pour cent (14,5 %) des ménages du fichier longitudinal se retrouvent dans cette situation.

Une variable dichotomique a été construite selon ces critères, et une régression logistique a été menée selon qu'un ménage fait partie de la sous-population d'intérêt ou pas (deux années sur trois

à 30 % ou plus, ou les trois années). La dimension temps ou durée est captée par cette variable et n'entre pas dans le modèle. Les variables indépendantes prennent leur valeur en début de période (2002)[58].

Le tableau 4.1 présente le sommaire général du modèle longitudinal persistant. Même si ce modèle teste une proposition différente du modèle longitudinal précédent, le réflexe premier est d'en comparer les résultats par variable. Ce modèle longitudinal recèle peu de surprises pour les six premières variables qui y apparaissent. Elles sont toutes fortement significatives, comme elles l'étaient dans le modèle longitudinal avec fichier de mesures répétées.

Tableau 4.1 Modèle – Persistant, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 (non significatif : seuil observé ≥ 0,05)

Variables	Seuils observés
Âge	0,0002
Sexe	0,0037
Princ. source revenu	0,0000
Genre de famille	0,0000
Activité principale	0,0151
Taille région résidence	0,0000
Déménagement	Non significatif
Niveau d'études	0,0023
Incapacité	Non significatif
Mode occupation	0,0122
Nbre pers. avec gains	Non significatif
Temps	_*

* Temps est implicite au modèle.

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004
Compilation : Société d'habitation du Québec.

La premier élément novateur intervient à la septième variable, soit la variable déménagement, qui devient non significative pour la première fois depuis le début de cette étude. Ce résultat semble indiquer que, si cette variable peut contribuer à expliquer un taux d'effort élevé l'année où ce déménagement s'effectue, celui-ci ne peut, par contre, étirer son influence au point d'expliquer deux années (ou trois) de taux d'effort élevés.

La variable niveau d'études obtient un seuil observé significatif, comparable à ce qui a été enregistré dans les modèles précédents. La variable incapacité demeure non significative.

Le deuxième élément novateur du tableau 4.1 se rapporte à la variable mode d'occupation, qui devient significative pour la première fois dans l'étude, avec un seuil observé de 0,012. Contrairement à ce que l'on observait pour la variable déménagement, le fait d'être propriétaire ou locataire semble contribuer à l'explication d'un taux d'effort élevé persistant, et ce, bien sûr, en sus de ce qu'apportent les autres variables du modèle (dont la variable principale source de revenus).

58. Une caractéristique expliquant la persistance doit être antérieure ou à parité dans le temps.

Finalement, le tableau 4.1 indique que la variable nombre de personnes avec gains redevient non significative, elle qui se qualifiait de justesse dans le modèle longitudinal avec fichier de mesures répétées.

Le tableau 4.2 présente les rapports de cotes associés aux niveaux de chacune de ces variables par rapport à leur niveau de référence. Ce tableau précise, le cas échéant, où se trouvent le ou les contrastes à l'intérieur de la variable qui fait en sorte que celle-ci est significative.

Tableau 4.2 Rapports de cotes, longitudinal « persistant » (2 ans ou 3 avec taux effort ≥ 30 %)

Variables	Taux d'effort de 30 % et plus pendant 2 ou 3 ans contre 30 % et plus pendant 1 an ou 0 an	
	Rapport de cotes	Seuils observés (Beta)
Âge		
De 15 à 29 ans	2,39	0,0033
De 30 à 44 ans	3,36	0,0000
De 45 à 64 ans	2,43	0,0000
65 ans et plus	1,00	-
Sexe		
Homme	0,64	0,0037
Femme	1,00	-
Principale source de revenus du ménage (4 catégories)		
Salaires et traitements	0,67	0,2086
Travail autonome	6,85	0,0000
Transferts gouvernementaux, autres revenus	3,67	0,0000
Pensions retraite et autres placements	1,00	-
Nombre de membres de la famille économique âgés de 16 ans et plus avec gains (3 catégories)		
Aucun membre		*n.s.*
Un membre		*n.s.*
Deux membres ou plus	1,00	-
Mode d'occupation		
Propriétaire	0,66	0,0120
Locataire	1,00	-
Genre de familles (5 catégories)		
Personne seule	2,71	0,0000
Couple sans enfants	1,00	-
Couple avec enfants	0,73	0,2200
Famille monoparentale	2,06	0,0170
Autres types	0,84	0,5750
Activité principale du soutien principal du ménage (4 catégories)		
Travaille	0,58	0,0320
Va à l'école	1,39	0,4350
Retraite, Codes réservés*	1,00	-
Autres	1,00	0,9890

* Regroupe les codes traitement provisoire, ne sait pas, refus et non applicable.

Tableau 4.2 Rapports de cotes, longitudinal « persistant » (2 ans ou 3 avec taux effort ≥ 30 %) (suite)

Variables	Taux d'effort de 30 % et plus pendant 2 ou 3 ans contre 30 % et plus pendant 1 an ou 0 an	
	Rapport de cotes	Seuils observés (Beta)
Taille de la région de résidence (5 catégories)		
Rurale	0,47	0,0000
Urbaine : 0 à 29 999 habitants	0,52	0,0002
Urbaine : 30 000 à 99 999 habitants	0,60	0,0070
Urbaine : 100 000 à 499 999 habitants	0,79	0,2355
Urbaine : 500 000 habitants et plus	1,00	-
Indicateur – État de l'incapacité pour l'année de référence		
Incapacité		*n.s.*
Non-incapacité	1,00	-
Codes réservés		*n.s.*
Statut de déménagement (3 catégories)		
Déménagement du ménage		*n.s.*
Déménagement du ménage après une séparation		*n.s.*
Ménage non déménagé, nouveau membre	1,00	-
Plus haut niveau d'études du soutien principal du ménage		
Moins qu'un D.E.S.**	1,00	-
Dipl. études secondaires	0,75	0,1010
Postsecondaire non universitaire	0,70	0,0870
Dipl. ou certificat universitaire	0,33	0,0001
Codes réservés	1,00	0,9950

** Diplôme d'études secondaires.

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

La première remarque qu'invite le tableau 4.2 est qu'avec ce modèle longitudinal dit persistant il n'y a plus deux logits, mais un seul. Ce modèle ne teste pas la même hypothèse que le modèle longitudinal avec fichier de mesures répétées. Il est donc difficile de comparer les rapports de cotes obtenus entre les deux modèles. Si l'on s'y laisse tenter toutefois, il vaut probablement mieux utiliser ceux du logit II du modèle répété (comparaisons des taux d'effort de 30 % à 49,9 % contre ceux de moins de 30 %), qui représente la situation de la majorité des ménages avec un taux d'effort égal ou supérieur à 30 %.

Si l'on compare les résultats observés dans ce tableau à ceux du logit II du modèle longitudinal de mesures répétées (tableau 3.6), on constate que les rapports de cotes et leurs seuils observés sont très similaires entre les deux modèles pour les trois premières variables : âge, sexe et principale source de revenus.

La variable nombre de membres de la famille de 16 ans ou plus avec des gains est non significative. Ses rapports de cotes ne sont pas présentés.

Avec le modèle persistant, la variable mode d'occupation connaît un revirement et le rapport de cotes indique que les ménages propriétaires et locataires se distinguent significativement. Les premiers courent moins de risques que les locataires de se situer dans des taux d'effort de 30 % et plus pendant deux ans ou trois ans.

La variable genre de famille produit des résultats similaires à ceux du logit II du modèle longitudinal avec fichier de mesures répétées. Deux types de ménages, les personnes seules et les familles monoparentales, se distinguent significativement du niveau de référence (les couples sans enfants).

Le seuil observé général de la variable activité principale annonçait le *statu quo* (la variable est significative), mais les rapports de cotes indiquent une nuance de taille : c'est l'activité principale de travailler qui se distingue significativement du niveau de référence (regroupant les personnes à la retraite et les codes réservés), alors que fréquenter l'école devient non significatif. Peut-être que de fréquenter l'école, tout comme l'activité de déménager pour la variable statut de déménagement, représente une activité transitoire qui ne peut expliquer une durée de deux ou trois ans dans les taux d'effort élevés. Chose certaine, la variation entre les niveaux de cette variable est très atténuée pour ce modèle. Le rapport de cotes pour le niveau « travaille » n'est significatif qu'à 0,032.

La variable taille de la région de résidence produit un seuil observé et des rapports de cotes très stables, très conformes à ce qu'on a vu précédemment. Les trois premiers niveaux se distinguent significativement du niveau de référence (urbain : 500 000 habitants et plus) : ce sont les ménages des régions rurales, des régions urbaines de 0 à 29 999 habitants, et des régions urbaines de 30 000 à 99 999 habitants qui affichent des probabilités inférieures de connaître un taux d'effort de 30 % et plus pendant deux ou trois ans.

Comme il a été mentionné précédemment, la variable statut de déménagement est non significative. Il n'y a donc pas lieu de s'intéresser à ses rapports de cotes. Finalement, la variable plus haut niveau d'études est significative, et résolument conforme à ses résultats antérieurs : seul le niveau diplôme ou certificat universitaire se distingue significativement de son niveau de référence (le niveau moins qu'un diplôme d'études secondaires).

4.2 MODÈLE DE TAUX D'EFFORT ÉLEVÉ PAR TRAJECTOIRE

L'utilisation d'un modèle longitudinal nommé « par trajectoire » résulte d'une collaboration de la statisticienne du Centre interuniversitaire québécois de statistiques sociales (CIQSS)[59]. Une courte présentation technique de ces modèles est présentée à l'annexe 8.

L'encadré qui suit donne les grandes lignes du processus. Il a pour seul objectif la compréhension du tableau 4.3 (prochaine page).

Les modèles de trajectoires

Les modèles de trajectoires découlent d'une famille de modèles en statistiques que l'on nomme modèles de mélange fini[60]. Un modèle de mélange fini s'intéresse d'abord aux valeurs des variables dépendantes. Dans ce cas-ci, elles sont établies à 1, 2, ou 3, selon le niveau des taux d'effort (1 = moins de 30 %, 2 = entre 30 % et 49,9 %, 3 = 50 % et plus). Le modèle de mélange fini modélise de façon longitudinale la variable mesurant ce niveau.

Les corrélations des valeurs de ces variables entre chaque année de la période d'observation sont établies. D'après la nature des données, le chercheur précise la fonction de distribution

59. M^me^ Danielle Forest. M^me^ Forest a travaillé avec le module « moc » du logiciel R.
60. Les modèles de classes latentes font aussi partie de cette famille de modèles.

qu'il entend appliquer aux observations. Dans le cas qui nous intéresse, les variables dépendantes ont trois niveaux. La fonction de distribution multinomiale a donc été choisie[61].

Le maximum de vraisemblance du modèle de mélange fini permet d'estimer la variable latente, laquelle permet à son tour d'établir le nombre de groupes homogènes que le modèle de mélange fini reconnaît dans les données. Le chercheur expérimente diverses solutions, et se fonde sur les indices d'ajustement pour effectuer un choix.

Après avoir défini le nombre de groupes optimal, la fonction de vraisemblance établit les probabilités *a posteriori* de chaque observation d'appartenir à chacun de ces groupes homogènes.

Dans le modèle retenu, la probabilité générale d'appartenir au groupe 1 est de 0,85, et la probabilité d'appartenir au groupe 2 est de 0,15. Cette probabilité est essentiellement une moyenne. Les valeurs moyennes pondérées du score de chaque groupe sont les suivantes :

Tableau 4.3 Score moyen de la variable dépendante pour les groupes issus de la fonction de mélange

	Période 1 (2002)	Période 2 (2003)	Période 3 (2004)
Groupe 1	1,07	1,03	1,06
Groupe 2	2,02	2,19	2,12

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

On relève donc que le groupe 1 se tient près du score moyen de 1,05 (1 = taux d'effort de moins de 30 %), et le groupe 2, du score moyen de 2,1 (2 = taux d'effort de 30 % à 49,9 %).

Comme il a été mentionné précédemment, à la suite de la définition des groupes, le modèle des trajectoires établit au niveau de chaque ménage sa probabilité *a posteriori* d'appartenir au groupe 1 et sa probabilité *a posteriori* d'appartenir au groupe 2. Les probabilités moyennes évoquées antérieurement (0,85 et 0,15) ne rendent pas compte de la différenciation (entropie) très forte au niveau des observations qui a été observée. Si l'on s'arrête au groupe 1, par exemple, plus des trois quarts des observations avaient des probabilités supérieures à 0,99 d'y appartenir[62].

Une fois ces probabilités ou poids établis en relation à la variable latente, les variables indépendantes (à leur valeur de 2002 – ou période un) sont introduites au fichier, et une analyse de régression logistique binomiale compare les caractéristiques des ménages selon leur poids relatif au groupe 1 contre leur poids relatif au groupe 2. Le tableau 4.4 rapporte les résultats généraux de ce modèle de régression logistique.

61. Des essais ont aussi été faits avec les variables dépendantes codées à deux niveaux (rapports d'abordabilité de moins de 30 %, contre 30 % et plus) et la distribution binomiale.
62. En corollaire, ils avaient donc des probabilités égales ou inférieures à 0,01 d'appartenir au groupe 2.

Tableau 4.4 Modèle – Trajectoire, seuils observés des variables retenues pour l'analyse du fichier longitudinal 2002-2004 (non significatif : seuil observé ≥ 0,05)

Variables	Seuils observés
Âge	0,0000
Sexe	0,0019
Princ. source revenu	0,0000
Genre de famille	0,0000
Activité principale	0,0058
Taille région résidence	0,0000
Déménagement	Non significatif
Niveau d'études	0,0004
Incapacité	Non significatif
Mode occupation	0,0172
Nbre pers. avec gains	Non significatif
Temps	_*

* Temps est implicite au modèle.

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

Les seuils observés de cet ensemble de variables indépendantes pour le modèle par trajectoire sont tous très proches de ceux du modèle longitudinal persistant (voir le tableau 4.1). Les six premières variables sont fortement significatives, comme elles l'étaient avec le modèle persistant. Elles le sont même davantage pour au moins deux de ces variables, soit activité principale (seuil de 0,006 plutôt que de 0,015) et niveau d'études (significative à 0,0004 plutôt qu'à 0,0023). La variable incapacité reste non significative, tout comme celle du nombre de personnes de 16 ans ou plus avec des gains. Ces deux variables ne semblent donc pas de nature à expliquer ni les taux d'effort élevés courants ni la persistance ou la durée dans les taux d'effort élevés. Par ailleurs, la variable déménagement demeure non significative, elle qui l'était pour la première fois avec le modèle longitudinal persistant. Enfin, le modèle par trajectoire confirme le résultat du modèle persistant en ce qui concerne la variable mode d'occupation, qui est significative cette fois-ci à 0,017. Elle semble donc avoir un effet en propre par rapport à la durée dans les taux d'effort élevés.

Nous constatons donc une très grande similitude des seuils observés des variables indépendantes de ce modèle par trajectoire avec ceux du modèle persistant. Comme précédemment, le tableau sommaire des seuils observés des variables indépendantes de la régression logistique peut être suivi du tableau des rapports de cotes à l'intérieur des variables, qui indiquent où se trouvent les contrastes intéressants.

Tableau 4.5 Rapports de cotes, longitudinal 2002-2004 - Trajectoires

Variables	Groupe 2 comparé au Groupe 1	
	Rapport de cotes	Seuils observés (Beta)
Âge		
De 15 à 29 ans	2,50	0,0000
De 30 à 44 ans	3,13	0,0000
De 45 à 64 ans	2,38	0,0000
65 ans et plus	1,00	-
Sexe		
Homme	0,65	0,0000
Femme	1,00	-
Principale source de revenus du ménage (4 catégories)		
Salaires et traitements	0,78	0,3700
Travail autonome	7,14	0,0000
Transferts gouvernementaux, autres revenus	3,85	0,0000
Pensions retraite et autres placements	1,00	-
Nombre de membres de la famille économique âgés de 16 ans et plus avec gains (3 catégories)		
Aucun membre		*n.s.*
Un membre		*n.s.*
Deux membres ou plus	1,00	-
Mode d'occupation		
Propriétaire	0,70	0,0200
Locataire	1,00	-
Genre de familles (5 catégories)		
Personne seule	2,63	0,0000
Couple sans enfants	1,00	-
Couple avec enfants	0,77	0,2300
Famille monoparentale	2,04	0,0100
Autres types	0,89	0,6700
Activité principale du soutien principal du ménage (4 catégories)		
Travaille	0,57	0,0100
Va à l'école	1,39	0,3800
Retraite, codes réservés*	1,00	-
Autres	1,01	0,9800
* Regroupe les codes traitement provisoire, ne sait pas, refus et non applicable.		
Taille de la région de résidence (5 catégories)		
Rurale	0,47	0,0000
Urbaine : 0 à 29 999 habitants	0,53	0,0000
Urbaine : 30 000 à 99 999 habitants	0,61	0,0000
Urbaine : 100 000 à 499 999 habitants	0,80	0,2000
Urbaine : 500 000 habitants et plus	1,00	-
Indicateur – État de l'incapacité pour l'année de référence		
Incapacité		*n.s.*
Non-incapacité	1,00	-
Codes réservés		*n.s.*

Tableau 4.5 Rapports de cotes, longitudinal 2002-2004 - Trajectoires (suite)

Variables	Groupe 2 comparé au Groupe 1	
	Rapport de cotes	Seuils observés (Beta)
Statut de déménagement (3 catégories)		
Déménagement du ménage		*n.s.*
Déménagement du ménage après une séparation		*n.s.*
Ménage non déménagé, nouveau membre	1,00	-
Plus haut niveau d'études du soutien principal du ménage		
Moins qu'un D.E.S.**	1,00	-
Dipl. études secondaires	0,74	0,0500
Postsecondaire non universitaire	0,69	0,0500
Dipl. ou certificat universitaire	0,33	0,0000
Codes réservés	1,02	0,9500

** Diplôme d'études secondaires.

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004. Compilation : Société d'habitation du Québec.

Tout comme le modèle longitudinal persistant, le modèle par trajectoire compare deux groupes (groupes 1 et 2) et ne crée donc qu'un logit. Le groupe 2 rassemble des ménages pour lesquels le modèle prédit des scores plus élevés de taux d'effort[63], alors que le groupe 1 rassemble des ménages pour lesquels le modèle prédit des taux d'effort acceptables. Ce dernier groupe étant placé au dénominateur du logit, les rapports de cotes sont établis de la même façon que précédemment, donc comparables à ceux du tableau 4.2, par exemple.

En ce qui concerne la variable âge, c'est pratiquement le *statu quo*. Les rapports de cotes bougent très peu, et tous les seuils observés sont significatifs à <0,0001. Même chose pour la variable sexe. Pour la variable principale source de revenus, encore une fois ne sont significatifs que les niveaux travail autonome et transferts gouvernementaux et autres revenus (par rapport au niveau de référence, soit les pensions de retraite et autres placements).

Comme nous l'avons mentionné précédemment, la variable nombre de personnes de 16 ans ou plus avec gains reste non significative.

La variable mode d'occupation produit un seuil observé très semblable, soit 0,02, comparativement à 0,012 pour le modèle persistant. Dans ce cas-ci, le rapport de cotes nous dit que les ménages propriétaires sont moins susceptibles de se voir classer dans le groupe 2, alors que, dans le modèle longitudinal persistant, ils étaient moins susceptibles d'avoir connu deux années ou trois de taux d'effort égal ou supérieur à 30 %.

Le résultat est similaire pour la variable genre de famille. Seuls les niveaux personnes seules et famille monoparentale se distinguent, avec des rapports de cotes et des seuils observés presque équivalents à ceux du modèle persistant.

L'activité principale du soutien principal indique, tout comme le modèle persistant, que c'est le niveau « travaille » qui se distingue significativement du niveau de référence. Le niveau « va à l'école » est de nouveau non significatif.

63. Leur score moyen varie entre 2,0 et 2,2 selon l'année. Le score de 2 est associé à des taux d'effort de 30 % à 49,9 % (voir le tableau 4.3).

Il y a aussi identité des résultats pour la variable taille de la région de résidence. Seul le niveau urbain (100 000 à 499 999 habitants) ne se distingue pas du niveau de référence.

Finalement, on observe des rapports de cotes très semblables pour la variable du plus haut niveau d'études. Ce qui change, ce sont les seuils observés de deux niveaux, les niveaux diplôme d'études secondaires et postsecondaire non universitaire. À un seuil observé de 0,05 dans le modèle longitudinal par trajectoire, ils se situent au tout début de la zone de non-significativité, alors qu'ils y étaient davantage engagés (à 0,10 et 0,09) dans le modèle persistant.

On constate donc que les deux modèles longitudinaux avec la dimension temps inhérente (ou implicite) produisent des résultats tout à fait semblables. Dans le premier de ces modèles longitudinaux, c'est par décision du chercheur que le regroupement des cheminements des ménages s'est fait. Tous les cheminements comportant deux ou trois années avec des taux d'effort égaux ou supérieurs à 30 % furent opposés aux cheminements ne comportant qu'une ou zéro année avec des taux d'effort égaux ou supérieurs à 30 %. En pratique, l'univers de ce dernier regroupement est celui des vecteurs [1 1 1], [1 1 2], [1 1 3], [1 2 1], [1 3 1], [2 1 1] et [3 1 1], donc sept vecteurs en tout opposés aux vingt autres.

Dans le deuxième de ces modèles longitudinaux, ce sont les corrélations entre les taux d'effort (les variables dépendantes) qui donnent naissance à une variable latente. L'intervention de l'analyste est davantage circonscrite. Ainsi, il doit trouver le nombre optimal de groupes pour rendre compte des trajectoires des ménages : dans cette étude, ce furent deux groupes. Le modèle par trajectoire établit les probabilités de chaque ménage d'appartenir à l'un et l'autre de ces groupes.

Pour vérifier comment les pondérations *a posteriori* du modèle par trajectoire se sont trouvées à fonctionner, nous avons regroupé les ménages selon leur pondération dominante (poids > 0,5)[64]. Le tableau 4.6 présente les résultats de cet exercice.

Tableau 4.6 Regroupement des vecteurs dans le modèle longitudinal par trajectoire

Vecteurs des ménages du Groupe 1	Vecteurs des ménages du Groupe 2
Probabilité (Groupe 1) > 0,5	Probabilité (Groupe 2) > 0,5
111	122
112	123
113	132
121	133
131	212
211	213
311	221
	222
	223
	231
	232
	233
	312
	313
	321
	322

64. Tout comme une probabilité, le poids maximal d'appartenir à un groupe est de 1,0 .

Tableau 4.6 Regroupement des vecteurs dans le modèle longitudinal par trajectoire (suite)

Vecteurs des ménages du Groupe 1	Vecteurs des ménages du Groupe 2
Probabilité (Groupe 1) > 0,5	Probabilité (Groupe 2) > 0,5
	323
	331
	332
	333

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

Comme on peut le constater, le modèle par trajectoire a rassemblé autour de la probabilité dominante d'appartenir au groupe 1 les ménages des mêmes vecteurs qui, dans le modèle persistant, avaient été considérés comme n'ayant pas connu de taux d'effort élevés. Les ménages de ces sept vecteurs se caractérisent tous par le fait qu'ils ont connu 0 ou 1 période seulement avec un rapport d'abordabilité égal ou supérieur à 30 % (sur les trois périodes d'observation retenues). Dans la formation du groupe 1, le modèle par trajectoire ne fait donc pas de distinction entre les niveaux 2 (30 % à 49,9 %) et 3 (50 % et plus) de la variable dépendante.

CONCLUSION

Ce chapitre a traité de modèles longitudinaux avec la dimension temps inhérente. Dans un de ces modèles (le modèle persistant), le regroupement des ménages a été effectué sur décision du chercheur. Les ménages considérés à taux d'effort élevés sont ceux qui ont connu deux ou trois périodes (années) avec un rapport d'abordabilité égal ou supérieur à 30 %. La distinction entre fardeau modéré (30 % à 49,9 %) et fardeau élevé (50 % et plus) est estompée, et, dans le cas de ceux qui ont connu deux périodes de taux élevé (et non trois), ces périodes peuvent être contiguës ou disjointes.

Dans l'autre modèle (le modèle par trajectoire), le regroupement des ménages a été effectué davantage en fonction de processus et de critères mathématiques. Dans les faits, ce modèle a produit un regroupement des ménages très similaire au modèle persistant, en autant que l'on ne tienne compte que de la « probabilité majoritaire » de chaque ménage d'appartenir à un groupe ou à l'autre. Le premier facteur de ressemblance est que, même si la variable dépendante comporte trois niveaux (taux d'effort de moins de 30 %, etc.), le modèle par trajectoire a estimé que la formation de deux groupes était ce qui s'adaptait le mieux aux données. Deux groupes aussi avaient été créés avec le modèle persistant.

Le deuxième facteur de ressemblance est que le modèle par trajectoire a regroupé les ménages selon le nombre de périodes où ils ont connu des taux d'effort acceptables (moins de 30 %) ou élevés (30 % et plus). Le groupe 1 a réuni tous les ménages ayant connu zéro ou une période seulement de taux d'effort élevé; le groupe 2 a réuni tous les ménages ayant connu deux ou trois périodes de taux d'effort élevé. Le modèle n'a pas tenu compte de critères supplémentaires (emplacement des périodes de taux d'effort élevés, distinction entre les niveaux 2 et 3 de taux d'effort, etc.).

Ces facteurs de ressemblance ont produit des résultats suffisamment convergents qu'il n'y a pas lieu, en conclusion, de ressasser les différences minimes de seuils observés et de rapports de

cotes entre les deux modèles à dimension temps inhérente. Nous allons plutôt faire un survol des résultats obtenus avec le modèle longitudinal persistant, qui est plus simple de présentation.

Parmi les variables prises en considération, celles qui contribuent à expliquer la durée dans les taux d'effort élevés persistants sont l'âge, le sexe, la principale source de revenu, le genre de famille, l'activité principale du soutien principal, la taille de la région de résidence, le niveau d'études et le mode d'occupation. La significativité de cette dernière variable constitue l'élément nouveau par rapport aux résultats des analyses antérieures (transversales et longitudinale avec fichier de mesures répétées).

Les variables écartées sont le statut de déménagement, l'indicateur d'incapacité et le nombre de membres de la famille économique de 16 ans ou plus avec gains. De ces variables écartées, la non-significativité de statut de déménagement constitue un revirement par rapport aux résultats antérieurs.

En ce qui a trait aux rapports de cotes, c'est-à-dire les contrastes à l'intérieur des variables, le modèle longitudinal persistant soumet certains résultats novateurs. Ainsi, pour la variable âge, tous les groupes d'âge se différencient significativement du groupe de référence (les 65 ans et plus), mais c'est le groupe du milieu, les 30 à 44 ans, qui obtient le rapport de cotes le plus élevé (3,36) de connaître un taux d'effort de 30 % ou plus pendant deux années (ou trois). Pour la variable principale source de revenus, ce sont ceux qui déclarent le travail autonome qui obtiennent ce rapport de cotes le plus élevé (6,85), devant ceux qui déclarent les transferts gouvernementaux ou les autres revenus comme principale source (3,67).

Les rapports de cotes pour les variables genre de famille et plus haut niveau d'études du soutien principal reproduisent les résultats antérieurs. D'une part, parmi les genres de familles, ce sont les personnes seules qui se démarquent le plus, et, d'autre part, parmi les niveaux d'études, ce sont les détenteurs d'un diplôme ou d'un certificat universitaire qui en font autant.

La variable activité principale du soutien principal amène également un élément nouveau, du fait que les soutiens qui fréquentent l'école obtiennent un rapport de cotes non significatif. Le seul rapport de cotes significatif produit par cette variable est celui des soutiens qui travaillent (0,58), qui ont une probabilité inférieure de connaître un taux d'effort égal ou supérieur à 30 % pendant deux ans (ou trois).

Pour sa part, la variable taille de la région de résidence génère trois niveaux, plutôt que deux comme les modèles antérieurs généralement, qui sont significativement différents du niveau de référence, soit les niveaux régions rurales, régions urbaines de 0 à 29 999 habitants, et les régions urbaines de 30 000 à 99 999 habitants.

Finalement, le rapport de cotes significatif produit par la variable mode d'occupation indique que les ménages propriétaires sont moins susceptibles que les locataires de connaître un taux d'effort élevé qui persiste deux ou trois ans (rapport de cotes : 0,66). Cette variable a donc un effet par rapport à la durée dans les taux d'effort.

CONCLUSION GÉNÉRALE

L'objectif premier de cette étude était de mieux connaître les ménages à taux d'effort très élevés, avec, comme sous-objectif particulier, de savoir s'il s'agit de situations temporaires ou de situations qui se prolongent. La période d'observation de l'étude est de trois ans (2002 à 2004). Pour répondre à ces questions, nous avons travaillé avec les données de l'Enquête sur la dynamique du travail et du revenu (EDTR), soit, en premier lieu, les échantillons transversaux de ménages québécois de cette période et, ensuite, avec l'échantillon longitudinal correspondant.

Avant d'énoncer les conclusions auxquelles nous amène cette recherche, il convient de rappeler quelle population est à l'étude. Celle-ci a été définie comme étant la population des ménages résidant au Québec, ayant des revenus annuels égaux ou supérieurs à 6 500 $ par année, et des coûts d'habitation annuels inférieurs à 100 000 $[65]. Ces deux derniers critères d'inclusion ont supprimé environ 2 % des enregistrements québécois (non pondérés) du fichier transversal de l'EDTR de 2003 (avec lequel cet exercice a été effectué).

La question de mieux connaître les ménages à taux d'effort élevés sera d'abord couverte à l'aide des analyses transversales (chapitre 2) ou à l'aide de l'analyse longitudinale avec fichier à mesures répétées (chapitre 3). Ces analyses, en effet, ont en commun d'avoir partagé la population à l'étude en trois niveaux de taux d'effort. D'un côté, il y a tous les ménages dont le taux d'effort est à l'intérieur de la norme d'abordabilité (logement : moins de 30 % du revenu avant impôt). De l'autre côté, il a été possible de subdiviser les ménages qui sont hors norme en deux groupes : ceux dont les coûts d'habitation vont de 30 % à 49,9 % de leurs revenus (fardeau moyennement élevé) et ceux dont ces coûts comptent pour 50 % et plus de leurs revenus (fardeau très élevé). Ces derniers représentent ni plus ni moins le groupe plus névralgique ayant servi de déclencheur à cette étude.

Afin de mieux connaître ces ménages, il est peut-être plus commode de privilégier les résultats obtenus à l'aide de l'analyse longitudinale avec fichier de mesures répétées, car ce dernier superpose les données des ménages qui sont présents les trois années. Cela peut favoriser la stabilité des résultats. Il faut garder à l'esprit toutefois que sa taille échantillonnale est conséquemment de deux à trois fois plus élevée que chacun des fichiers transversaux, ce qui favorise la significativité[66].

Pour l'échantillon longitudinal, sur une base **annuelle**, ce sont environ 84 % des ménages de cet échantillon qui obtiennent un taux d'effort acceptable, c'est-à-dire inférieur à 30 %[67]. Seize pour cent des ménages obtiennent donc un rapport d'abordabilité supérieur à la norme. La répartition est d'environ deux tiers/un tiers entre ceux qui ont un rapport d'abordabilité avec fardeau moyennement élevé (coût d'habitation entre 30 % et 49,9 % du revenu : 11 % de l'échantillon) et avec fardeau très élevé (coût d'habitation de 50 % et plus du revenu : 5 % de l'échantillon). Ces données figurent à l'annexe 9[68].

En excluant la variable temps, il y onze variables qui entrent dans le modèle longitudinal avec fichier de mesures répétées (voir le tableau 3.4, p. 45). Les variables qui vont nous renseigner sur les taux d'effort élevés en général (égaux ou supérieurs à 30 %) sont l'âge, le sexe, la principale source de revenus, le genre de

65. En ce qui concerne les fichiers transversaux, les critères s'appliquent à l'année en considération. Pour le fichier longitudinal, l'ensemble des critères, dont la résidence au Québec, s'appliquent à l'année 2004.
66. D'autant plus que la corrélation longitudinale n'est pas prise en compte.
67. Ce pourcentage est légèrement supérieur à celui qui est obtenu avec les échantillons transversaux (en 2003, 81,5 % des ménages québécois obtenaient un taux d'effort acceptable).
68. Avec l'échantillon longitudinal, nous obtenons donc environ +2 % de ménages avec un taux d'effort acceptable, si l'on compare avec les résultats du tableau 1.2 (obtenus avec le fichier transversal de 2003 – voir p.18). Le fichier longitudinal pourrait privilégier les ménages qui sont plus stables dans leur composition, ce qui se répercute peut-être sur les taux d'effort.

familles, l'activité principale, la taille de la région de résidence, le statut de déménagement et le niveau d'études[69]. En effet, ces variables sont celles qui sont significatives dans ce modèle de fichier longitudinal.

L'effet général de ces variables a été traité dans la conclusion du chapitre 3, et nous n'y reviendrons pas. Il nous semble plus important, dans cette conclusion générale, de faire ressortir ce qui paraît propre aux ménages avec fardeau des coûts très élevés, le groupe plus névralgique, comme il a été mentionné plus tôt.

L'âge est l'une des variables qui permet de caractériser les ménages à taux d'effort très élevés, car chacun des groupes d'âge du modèle logit sous-jacent obtient des rapports de cotes supérieurs à ceux du logit des taux d'effort de 30 % à 49,9 % (voir le tableau 3.6, p. 46). Le groupe d'âge pour lequel l'écart entre les rapports de cotes est le plus grand est le groupe des 45 à 64 ans : la probabilité relative de ce groupe d'âge de figurer parmi les ménages avec fardeau très élevé est vraiment plus forte que celle de figurer parmi les ménages avec fardeau moyennement élevé (rapport de cotes, fardeau très élevé = 4,6; rapport de cotes, fardeau moyennement élevé = 2,1).

Par contre, la variable sexe est neutre à cet égard : les rapports de cotes sont égaux.

La variable principale source de revenus ne saurait caractériser les ménages à taux d'effort très élevés que pour le niveau des travailleurs autonomes, qui ont un rapport de cotes sensiblement plus élevé au logit des taux d'effort de 50 % et plus (rapport = 9,5) qu'au logit des taux d'effort de 30 % à 49,9 % (rapport = 6,1). Cependant, c'est un groupe de petite taille; la prudence reste de mise. Si l'on s'attarde à la catégorie des ménages qui tirent leur principale source de revenus des transferts gouvernementaux, catégorie à laquelle plusieurs ménages résidant en HLM vont appartenir, ces ménages obtiennent un rapport de cotes qui n'est que faiblement plus élevé au logit des taux d'effort égal ou supérieur à 50 %, comparativement au logit des taux de 30 % à 49,9 %. Les probabilités relatives de ces ménages de connaître un fardeau très élevé plutôt qu'un fardeau moyennement élevé ne semblent pas supérieures.

Dans la liste des variables significatives mentionnées plus haut, la variable genre de familles arrive à la suite. Parmi les quatre niveaux ou catégories qu'elle recouvre en plus du niveau de référence, seul le niveau personne seule permet de caractériser la population des ménages avec fardeau des coûts élevé. Parmi les ménages qui connaissent un taux d'effort de 50 % et plus, ceux qui sont composés de personnes seules obtiennent des probabilités supérieures d'y figurer que de figurer parmi les ménages qui connaissent un taux d'effort de 30 % à 49,9 %.

Un résultat similaire s'applique à la variable activité principale du soutien principal. Parmi ses quatre niveaux, seul le niveau de fréquenter l'école permet de caractériser les ménages avec fardeau élevé. Parmi les ménages avec fardeau très élevé, il y a plus de chances que l'on retrouve des ménages dont le soutien principal va à l'école, plutôt que de retrouver ces derniers parmi les ménages avec fardeau moyennement élevé.

La variable significative suivante est la taille de la région de résidence. Cette variable se compare à la variable âge, en ce sens que son influence par rapport aux taux d'effort de 50 % et plus est généralisée pour tous ses niveaux. En plus de son niveau de référence (les ménages des régions urbaines des 500 000 habitants et plus), cette variable compte quatre niveaux. Or ceux-ci produisent tous des rapports de cotes inférieurs pour la proposition de connaître des taux d'effort de 50 % et plus que ceux qu'ils produisent pour la proposition de connaître des taux d'effort entre 30 % et 49,9 % (fardeau moyennement élevé). Autrement dit, parmi les ménages avec fardeau élevé des coûts, ceux des régions rurales (par exemple) obtiennent une probabilité moindre d'y figurer que leur probabilité relative de figurer parmi la population des ménages avec fardeau modéré des coûts. La même chose s'applique aux ménages des régions urbaines de 0 à 29 999 habitants, aux

69. La variable nombre de personnes de 16 ans ou plus avec des gains a été exclue, car le seuil faiblement significatif obtenu faisait exception aux résultats transversaux, de même qu'aux autres résultats longitudinaux. Les autres variables non significatives sont le mode d'occupation et l'indicateur d'incapacité.

ménages des régions urbaines de 30 000 à 99 999 habitants, et aussi aux ménages des régions urbaines de 100 000 à 499 999 habitants[70].

La variable statut de déménagement provoque également un effet généralisé par rapport aux taux d'effort de 50 % et plus, en ce sens que cet effet se manifeste dans les deux niveaux autres que son niveau de référence. Parmi les ménages avec fardeau très élevé, tant les ménages qui ont déménagé à la suite d'une séparation que ceux qui ont déménagé pour toute autre raison obtiennent une probabilité relative supérieure (par rapport au niveau de référence) d'y figurer que celle qu'ils obtiennent parmi les ménages avec un fardeau des coûts moyennement élevé[71].

Finalement, il ne reste que la variable plus haut niveau d'études du soutien principal qui peut servir à caractériser la population des ménages à taux d'effort très élevé de 50 % et plus. Cette variable ne semble pas vraiment utile à cette caractérisation. En effet, parmi les ménages à taux d'effort très élevé, seuls ceux dont les membres détiennent un diplôme ou un certificat universitaire se différencient significativement du niveau de référence (le niveau « moins qu'un diplôme d'études secondaires »). Leur rapport de cotes est de 0,44. Or parmi les ménages avec fardeau des coûts moyennement élevé, ceux dont les membres détiennent un diplôme ou un certificat universitaire se différencient également de façon significative, et ce, avec un rapport de cotes de 0,40. Essentiellement, les probabilités de ces détenteurs de diplôme sont les mêmes, peu importe qu'il s'agisse de figurer parmi la population des ménages avec fardeau très élevé ou parmi la population des ménages avec fardeau moyennement élevé.

Parmi les huit variables significatives du modèle longitudinal avec mesures répétées, il y en a donc trois qui semblent avoir un effet généralisé sur l'ensemble de leurs niveaux relativement aux ménages avec fardeau des coûts très élevé : l'âge, la taille de la région de résidence et le statut de déménagement. Ces variables vont être associées à de plus fortes probabilités relatives[72], pour les ménages des niveaux pertinents, à figurer parmi les ménages avec fardeau des coûts très élevé que parmi les ménages avec fardeau des coûts moyennement élevé.

Deux autres variables ont un effet localisé relativement aux ménages avec fardeau des coûts très élevé : le genre de familles et l'activité principale du soutien. Seul un niveau de ces variables va engendrer des probabilités relatives plus élevées de figurer parmi la population des ménages avec fardeau des coûts très élevé plutôt que parmi la population des ménages avec fardeau moyennement élevé. Ce sont les niveaux personne seule (variable genre de famille) et va à l'école (activité principale).

La principale source de revenus aurait pu être classée parmi ces variables avec effet localisé, si les écarts de rapport de cotes pour le niveau travail autonome avaient été un peu plus accentués. À défaut, nous la classons parmi les variables neutres sur cette question[73].

Finalement, deux autres variables peuvent être qualifiées de neutres : le sexe du soutien principal et le plus haut niveau d'études. Il n'y a pas de différence sensible dans les rapports de cotes des niveaux pertinents entre la population des ménages avec fardeau élevé et celle avec fardeau modéré.

Donc, les ménages avec fardeau élevé ne se différencient pas de façon systématique des ménages avec fardeau modéré. Il n'y a que trois variables, sur onze finalement (en incluant les variables non significatives), qui pointent en ce sens. Deux autres pointent vers une différenciation localisée à un niveau de la variable, et non pas à l'ensemble de ceux-ci. En conclusion, le fait d'appartenir aux ménages à taux d'effort élevé (50 % et plus) peut se traduire par certaines répercussions sur les rapports de cotes des niveaux d'une variable, mais

70. Ces deux derniers niveaux n'étaient pas ressortis comme significatifs dans le modèle transversal de 2003.
71. Dans les modèles transversaux, cette variable n'était pas significative pour 2002.
72. Ou, dans le cas de la variable taille de la région de résidence, à de plus faibles probabilités relatives, étant donné que le niveau de référence de cette variable a été attribué au niveau ayant les probabilités les plus élevées de connaître un taux d'effort supérieur (pointage à l'inverse pour cette variable).
73. C'est ce que laissent clairement voir les résultats du modèle transversal de 2003.

pas de façon dramatique, ni par revirement, généralement parlant. Nous ne croyons pas que l'hypothèse générale voulant qu'il s'agisse de groupes à caractéristiques distinctes soit validée[74].

Nous avons mentionné plus tôt que les modèles transversaux et le modèle longitudinal avec fichier de mesures répétées étaient les modèles avec lesquels on pouvait traiter de la différenciation entre les taux d'effort de 30 % à 49,9 %, contre les taux de 50 % et plus. Cet aspect ayant été couvert, nous en arrivons maintenant à la deuxième question sur laquelle portait cette recherche : qu'en est-il de la durée dans les taux d'effort? Est-ce que ce sont toujours les mêmes ménages qui se retrouvent avec des taux d'effort élevés?

Le tableau 3.2 (p. 42) indiquait que les trois quarts des ménages dans la population cible étaient demeurés avec des taux d'effort acceptables durant les trois années de la période d'observation. Sur l'ensemble de celle-ci, les taux d'effort égaux ou supérieurs à 30 % concernent donc le quart des ménages. Que pouvons-nous conclure de la durée dans les taux d'effort élevés de ces ménages?

À titre factuel, parmi ce quart restant des ménages, 40 % n'ont connu qu'une période de taux d'effort égal ou supérieur à 30 %[75], et 60 % ont connu deux ou trois de ces périodes[76]. Ces derniers forment donc la majorité, signe d'une certaine concentration[77].

Les ménages qui ont connu deux ou trois périodes avec taux d'effort égal ou supérieur à 30 % ont donné naissance au modèle longitudinal, dit persistant. Ce modèle compare ce groupe de ménages à tous les autres (zéro ou une période de taux d'effort égal ou supérieur à 30 %). Parmi les onze variables en entrée du modèle, le modèle de régression logistique comparant ces deux groupes donne comme significatifs l'âge, le sexe, la principale source de revenus, le genre de famille, l'activité principale, la taille de la région de résidence, le niveau d'études et le mode d'occupation. La significativité de cette dernière variable constitue l'élément neuf par rapport aux résultats précédents de cette étude. Sur une période d'une année, cette variable ne peut expliquer un taux d'effort élevé. Par contre, sur une perspective plus longue, si le taux d'effort élevé du ménage se prolonge, elle peut en éclairer la situation.

L'inverse semble s'appliquer à la variable statut de déménagement. Elle n'est pas significative pour le modèle longitudinal persistant, mais était significative dans les modèles antérieurs. Il semble donc qu'elle peut contribuer à l'explication d'un taux d'effort élevé l'année où le déménagement a lieu, mais elle n'interviendrait pas dans le cas d'un taux d'effort élevé qui se prolonge[78]. Ces deux variables (mode d'occupation et statut de déménagement) nous rappellent donc que, malgré leur similitude, ce modèle persistant et le modèle précédent ne testent pas la même chose. Un résultat à la manière transversale n'est pas nécessairement garant d'un résultat longitudinal similaire. En ce sens également, les rapports de cotes obtenus pour le modèle persistant (tableau 4.2, p. 53) ne sont pas directement comparables à ceux du modèle longitudinal avec fichier de mesures répétées (tableau 3.6).

Une différence qu'il importe de rappeler concerne la variable activité principale du soutien principal. Les modèles transversaux et le modèle longitudinal avec fichier de mesures répétées nous avaient habitué à voir le niveau de fréquenter l'école comme le niveau qui se démarquait des autres. Cependant, avec le modèle longitudinal persistant, ce niveau devient non significatif. Puisque la variable reste significative, le niveau qui le remplace est le niveau travaille, significatif au seuil de 0,032 (pour un rapport de cotes de 0,58). Donc, pour

74. Dans cette conclusion, nous sommes également encouragés par le fait que le modèle longitudinal par trajectoire ne fait pas de distinction entre les niveaux 2 et 3 de sa variable dépendante (ces niveaux correspondent aux taux d'effort avec fardeau moyennement élevé et avec fardeau très élevé).
75. 534 ménages sur 1 234. Calculé à partir du tableau 3.2.
76. 700 ménages sur 1 234.
77. Il est très probable que, sous une période d'observation plus longue que trois ans, disons la pleine période de six ans, la proportion des ménages qui ne connaissent qu'une année de taux d'effort élevé diminuerait.
78. Les modèles longitudinaux à dimension temps inhérente ne prennent en considération que la valeur d'une seule période de la variable indépendante. Par contre, si un ménage connaissait deux périodes avec déménagement, le résultat transversal risquerait de se reproduire.

cette variable, le niveau va à l'école ne peut expliquer une durée prolongée dans les taux d'effort égaux ou supérieurs à 30 %. Par contre le niveau travaille est associé à une probabilité plus faible de connaître une durée prolongée dans ces taux d'efforts.

Pour les autres variables, les niveaux qui se démarquent dans le modèle persistant sont généralement les mêmes que dans le modèle longitudinal avec fichier de mesures répétées. Les variations sont minces.

En somme, parmi les onze variables entrées dans le modèle à mesures répétées et dans le modèle persistant, il y en a deux que l'on peut considérer comme non significatives : l'état de l'incapacité et le nombre de membres de la famille économique avec des gains[79].

Il y a en a six qui paraissent jouer sur les deux tableaux, c'est-à-dire le taux d'effort actuel du ménage, et la durée dans ce taux d'effort. Ce sont l'âge, le sexe, la principale source de revenu, le genre de famille, la taille de la région de résidence et le niveau d'études.

Une autre variable joue également sur les deux tableaux, mais avec une nuance importante : l'activité principale. Le niveau va à l'école est clairement associé au taux d'effort actuel du ménage, mais n'est pas associé à la durée dans le taux d'effort[80].

Finalement, deux variables ne sont significatives que pour l'une de nos interrogations. Le statut de déménagement est associé au taux d'effort actuel du ménage, mais pas à la durée, tandis que c'est l'inverse pour le mode d'occupation. Cette variable ne semble pas influencer le taux d'effort actuel du ménage, compte tenu des autres variables dans le modèle, mais intervient sur la persistance dans ce taux d'effort.

PERSPECTIVES

Les conclusions de cette étude ne doivent pas être sorties de leur contexte. Celui-ci est donné par les questions auxquelles nous avons tenté de répondre, et qui ont été rappelées dans cette conclusion : qu'est-ce qui explique un fardeau des coûts moyennement élevé ou très élevé pour un ménage dans une année donnée, et qu'en est-il de la durée dans ce fardeau des coûts?

L'analyse repose également entièrement sur une approche multivariée. On effectue une analyse à plusieurs variables pour dégager du groupe celles qui sont les plus importantes. À ce titre, la conclusion qui risque probablement de susciter le plus d'interrogations est celle concernant la variable mode d'occupation, qui ne semble pas significative pour expliquer le taux d'effort courant des ménages (elle est significative toutefois pour expliquer la persistance dans les taux d'effort élevés).

Cette conclusion est tirée d'un contexte multivarié particulier. Les variables insérées simultanément dans le modèle réduisent le pouvoir explicatif de cette variable. Dans des contextes non multivariés, le rôle tenu par le mode d'occupation semble primordial. Ainsi, l'étude de Luffman, dès son premier paragraphe, fait valoir que :

> Toutefois, ces coûts (d'habitation) diffèrent de façon importante chez les propriétaires et les locataires. Un peu plus de 20 % du budget des ménages est consacré aux coûts d'habitation chez les locataires, par rapport à seulement 13 % chez les propriétaires[81].

Dans son contexte propre, cette observation est indéniable. Il est regrettable toutefois que l'étude de Luffman ait omis d'inclure la variable mode d'occupation dans sa section qui ouvre sur une analyse multivariée[82]. En effet, cette section de l'étude porte uniquement sur les ménages locataires. L'auteure a sans doute considéré

79. Le modèle longitudinal avec mesures répétées est le seul qui donne, en fait, cette variable comme significative, mais à un seuil de 0,043. Nous avons considéré accidentel ce résultat, sans y donner suite.
80. Ce niveau est sans doute pénalisé par son petit n. La variable reste significative dans le modèle persistant en raison du niveau travaille, associé à une probabilité moindre de connaître un taux d'effort élevé persistant.
81. Luffman (2006), *op. cit.*, p. 17.
82. *Ibid.*, p. 24-25.

que propriétaires et locataires sont tellement différents qu'ils ne peuvent être mis dans le même modèle. Ce choix rend donc non directement comparables l'analyse multivariée de l'étude de Luffman et celle-ci puisque les populations à l'étude ne sont pas les mêmes.

Les résultats de cette étude-ci indiquent qu'il ne faut pas tenir pour acquise la variable mode d'occupation. Il nous semble préférable de l'inclure explicitement dans les modèles multivariés.

Un autre choix qui a été fait, et qui mérite considération, est celui d'inclure non pas le revenu du ménage dans le modèle (puisqu'il est présent dans la variable dépendante), mais une variable quand même reliée d'assez près à celui-ci, par exemple la principale source de revenu. L'étude SCHL (2005) et l'étude de Luffman l'ont incluse, et nous avons reproduit ce choix. Puisque « le revenu des ménages est un facteur clé[83] » dans l'étude des taux d'effort, il faut une variable apparentée satisfaisante si l'on veut éviter que l'importance d'autres variables qui pourraient être reliées également au revenu gonfle indûment. L'annexe 6 donne précisément un exemple d'une variable reliée d'assez près aux niveaux de revenu qui a pu jouer ce rôle : la principale source de revenus.

Un autre résultat de cette recherche sur lequel nous aimerions attirer l'attention est le rôle de la variable taille de la région de résidence. Se limiter aux tableaux bivariés entre le taux d'effort et la taille de la région de résidence (tableaux 1.4 et 2.1) n'aurait pas rendu justice à cette variable. Dans les modèles, son seuil observé a toujours été systématiquement très bas (<0,0001). Cette variable est pertinente pour comprendre à la fois la situation actuelle d'un ménage et la durée dans son taux d'effort. Elle nous dit, en gros, que les ménages des régions rurales et des régions urbaines peu populeuses sont moins susceptibles que les ménages des régions très peuplées de connaître un taux d'effort égal ou supérieur à 30 %.

Il serait sans doute prétentieux de présenter ce résultat comme une grande innovation, puisque la SCHL a toujours accordé une forte importance aux différences géographiques. Cependant elle a parfois tendance à mettre l'accent sur les « différences entre les grandes villes du Canada[84] » ou les différences entre les provinces et les territoires[85]. Dans le contexte de cette étude-ci, l'originalité de la variable taille de la région de résidence est de mettre en évidence l'aspect intrarégional, selon le degré de densité des localités et des territoires qui composent une province donnée.

Le genre de famille économique est une variable dont ont fait usage tant l'étude de Luffman que l'étude de la SCHL. L'étude de Luffman y fait référence au tableau 3 (p. 22 de l'étude). Dans le texte d'accompagnement du tableau et dans la conclusion, l'auteure a tendance à faire ressortir le rôle tenu par les personnes seules, tant les personnes âgées que les non âgées (voir p. 21). L'étude de la SCHL (2005), pour sa part, fait ressortir le rôle des familles monoparentales (d'abord) et des ménages non familiaux (les personnes seules, voir p. 6). Dans notre étude, les rapports de cotes ont eu tendance à démarquer la part tenue par les personnes seules. Les rapports de cotes des ménages monoparentaux, quant à eux, sont généralement significatifs, mais se classent le plus souvent au deuxième rang (par rapport au niveau de référence de la variable).

Une des réponses que l'étude de la SCHL (2005) fournissait à la question « Quels groupes sont susceptibles de consacrer 50 % ou plus de leur revenu au logement? » était les jeunes (voir SCHL, 2005, p. 5). Dans notre étude, les rapports de cotes associés à la variable âge indiquent généralement que ce sont les ménages dirigés par une personne de 65 ans ou plus qui sont en meilleure position. Les ménages dirigés par des jeunes, effectivement, sont plus susceptibles de connaître des taux d'effort élevés, mais leur situation relative n'est pas particulièrement plus déplorable que la situation des groupes des 30 à 44 ans ou des 45 à 64 ans[86].

---

83. *Ibid.*, p. 24.
84. SCHL (2005), « Le point en recherche », nº 8 révisé, *op. cit.*, p. 6. Tableaux aux pages suivantes.
85. SCHL (2004), « Le point en recherche », nº 3 révisé, *op. cit.*, tableaux de l'annexe, p. 6 et suivantes.
86. Le modèle transversal de l'année 2002 est celui pour lequel la situation des jeunes paraît la plus difficile, et ce, pour le logit des taux d'effort de 50 % et plus (contre les 0 % à 29,9 %). Leur rapport de cotes est à 7,4 , alors que les deux autres groupes d'âge sont à 4,3 et 4,2 .

Notre étude avait fait place à deux variables concernant l'état de santé (état de santé actuel et état de l'incapacité), qu'on associe souvent à des taux d'effort élevés. Ainsi, l'étude de Luffman relève (en mode bivarié) que 44 % des locataires dont le fardeau des coûts d'habitation était élevé « avaient une incapacité physique, par rapport à seulement 16 % des ménages sans problème d'abordabilité[87] ». Pourtant, ni l'une ni l'autre de ces variables ne s'est imposée dans les analyses multivariées menées. Il est à noter que le résultat concernant la variable état de l'incapacité est congruent avec le propre résultat du modèle multivarié pour ménages locataires de Luffman[88].

Enfin, notre étude met en évidence le rôle relatif aux taux d'effort élevés de deux situations particulières pour un soutien principal du ménage, soit celui d'aller à l'école et celui de déménager. La probabilité de taux d'effort plus élevé à cause de l'une ou l'autre de ces situations vaut pour l'année courante de celle-ci. Dans le cas particulier du statut de déménagement, cela vaut peu importe la raison (séparation ou autre). Parmi les autres raisons qui peuvent inciter à déménager, il n'est pas impensable que l'une de celles-ci soit de vouloir diminuer son taux d'effort (diminuer son coût de logement). Toutefois, les données de cette étude indiquent que l'aboutissement le plus usuel d'un déménagement est d'entraîner une augmentation, à tout le moins passagère, de ce taux d'effort.

Un dernier point à souligner concerne la variable principale source de revenus. Les autres études qui ont été recensées (Luffman, 2006, p. 21; SCHL, 2005, p. 4-5) ont eu tendance à relever que le fait de dépendre des transferts gouvernementaux comme principale source de revenus était un bon prédicteur de taux d'effort élevés. C'est le niveau de la variable qui est généralement mis en évidence. Notre étude indique toutefois que le fait d'être travailleur autonome est aussi susceptible d'entraîner des taux d'effort plus élevés. Avec le modèle longitudinal persistant, son rapport de cotes devient même supérieur à celui de transferts gouvernementaux (rapport de cotes de 6,7 et 3,9, respectivement). Nous croyons que les résultants observés pour le travail autonome devraient être validés par toute nouvelle étude qui se pencherait sur la question[89].

À ce sujet, il est tout à fait probable que de nouvelles études seront menées dès lors que les données de l'année 2007 deviendront disponibles, soit, normalement, au cours de 2009. À ce moment, les données pour le panel 4 deviendront complètes : la suite entière des six années de 2002 à 2007 sera achevée[90]. La dimension longitudinale pourra être approfondie.

87. Luffman, *op. cit.*, p. 21.
88. Voir son tableau 4, p. 24.
89. Dans le modèle de régression relatif aux ménages locataires de l'étude de Luffman, on constate également que le niveau travail autonome obtient un rapport de cotes significatif. Cependant, l'auteure, dans sa discussion (voir p. 25), minimise ce résultat parce que le rapport de cotes pour le travail autonome n'est significatif qu'avec la méthode du rapport d'abordabilité selon le revenu, mais pas avec la méthode concurrente (développée dans son étude) du rapport d'abordabilité selon les dépenses. Notons que, dans l'EDTR, les revenus de travail autonome englobent ceux du travail autonome agricole et ceux du travail autonome non agricole.
90. Rappelons que les informations sur les coûts d'habitation sont apparues dans l'EDTR en 2002. En plus du panel 4, nous disposerons également, en 2009, des trois premières années du panel 5 (2005-2007).

BIBLIOGRAPHIE

BROOKS, Bradley (2005). *Le chômage chronique : un profil statistique*, Statistique Canada, nº 11-621-MIF2005-031 au catalogue, coll. « Analyse en bref », Ottawa, 14 p.

FOREST, Danielle (2007). *Modèles de trajectoires.* Présentation sous forme de document Acrobat Reader faite à l'Université Laval le 28 mars 2007, 7 p.

GAMEROFF, Marc J., *Using the Proportional Odds Model for Health-Related Outcomes : Why, When, and How with various SAS Procedures*, Paper 205-30, SUGI 30, 8 p.

GOUVERNEMENT DU QUÉBEC. Ministère de l'Emploi et de la Solidarité sociale, *Rapport statistique sur les programmes d'assistance sociale*, juillet 2007, tableau 1, p. 1-2.

GOUVERNEMENT DU QUÉBEC. Office québécois de la langue française, *Le grand dictionnaire terminologique.* Voir http://www.olf.gouv.qc.ca/ressources/gdt.html.

LUFFMAN, Jacqueline, « Mesurer l'abordabilité du logement », revue *L'emploi et le revenu en perspective*, de Statistique Canada, nº 75-001-XIF au catalogue, nov. 2006, vol. 7, nº 11, p. 17-27

REA, Willa, et J. Yuen (2007). *La dynamique de l'abordabilité des logements : résultats provisoires du projet de recherche conjoint SCHL-Statistique Canada.* Présentation Powerpoint au Comité national de recherche sur le logement (CNRL), 1er mai (22 diapositives).

SCHL (2003). « Le point en recherche », série socioéconomique 03-017, série sur le logement selon les données du recensement de 2001 : nº 1, *Amélioration sur le plan de l'abordabilité du logement*, 10 p.

SCHL (2004). « Le point en recherche », série socioéconomique 04-007, série sur le logement selon les données du recensement de 2001 : nº 3 révisé, *Qualité, taille et abordabilité du logement canadien*, avril 2004, 10 p.

SCHL (2005). « Le point en recherche », série socioéconomique 05-004, série sur le logement selon les données du recensement de 2001 : nº 8 révisé, *Ménages qui consacrent 50 % ou plus de leur revenu au logement*, avril 2005, 12 p.

STATISTIQUE CANADA. *Dictionnaire du recensement de 2001*, mars 2002, nº 92-378-XIF01000, 394 p.

STATISTIQUE CANADA. *Enquête sur la dynamique du travail et du revenu : dictionnaire électronique des données.* Voir http://statcan.ca/francais/freepub/75F0026XIF/75F0026XIF2006000.htm.

STATISTIQUE CANADA. *Fichier des microdonnées à grande diffusion du Recensement de 2001. Fichier des ménages et du logement.*

STATISTIQUE CANADA. *Guide de l'utilisateur des microdonnées de l'Enquête sur la dynamique du travail et du revenu*, cat. nº 75M0001GPF, 1994, 3221 p.

STATISTIQUE CANADA. *Méthodologie de l'enquête sur la population active du Canada*, nº 71-526-XPB au catalogue, ministre de l'Industrie, 1998, 51 p., annexes A-D non paginées.

ANNEXES

ANNEXE 1

PRÉCISIONS RELATIVES À LA NORME DU RAPPORT DU COÛT D'HABITATION AU REVENU

Dans les constituantes du taux d'effort, on peut soumettre que la SCHL a défini avec précision ce que comprennent les coûts d'habitation. Pour tous les ménages, ils comprennent les frais d'électricité, du combustible, de l'eau et des autres services municipaux, auxquels s'ajoutent :

- pour les locataires : le loyer;
- pour les propriétaires : les paiements hypothécaires (principal et intérêts), les impôts fonciers (taxes municipales et scolaires) et, le cas échéant, les charges de copropriété.

Tous ces éléments se retrouvent aux questions H5 à H8 du formulaire long du recensement (version 2006). Ils sont rappelés dans plusieurs des publications de la SCHL, du groupe « Le point en recherche », série socioéconomique.

Par contre, les éléments inclus dans le dénominateur du taux d'effort – le revenu total avant impôt – ne sont pas précisés de façon aussi succincte et claire, du moins récemment. Il semble exister une certaine ambiguïté due au fait que l'on retrouve deux instruments, ou deux méthodes à vrai dire, pour établir le revenu.

D'une part, il y a l'instrument en quelque sorte original (du point de vue de la SCHL), qui est le formulaire long du recensement. Les catégories prévues en matière de revenu se retrouvent à la question 52 du formulaire de 2006 : revenu de travail rémunéré, revenu de travail autonome, revenu provenant du gouvernement, et « autre revenu ».

D'autre part, il y a l'instrument que constitue le questionnaire de l'Enquête sur la dynamique du travail et du revenu (EDTR). Dans le *Guide de l'utilisateur des microdonnées de l'EDTR*, il est rapporté qu'à l'origine du moins « l'EDTR utilise le même questionnaire sur le revenu que l'Enquête sur les finances des consommateurs[91] ». Toutefois, puisque l'EDTR donne le choix au répondant entre répondre de lui-même au questionnaire sur le revenu ou autoriser la consultation de sa déclaration de revenus à l'Agence du revenu du Canada, il se pourrait que ce choix ait eu tendance à rapprocher les données recueillies par l'EDTR de celles de la déclaration de revenus[92].

Les données sur le revenu de l'EDTR seraient donc davantage influencées par l'instrument fiscal que par celui du questionnaire long du recensement. Or ces deux instruments ne concordent pas parfaitement. Beaucoup de divergences proviennent de la catégorie « autre revenu » du formulaire long du recensement.

En premier lieu, ce dernier utiliserait le concept du « revenu monétaire total ». Conséquemment, sa catégorie « autre revenu » exclut les gains ou pertes en capital, de même que les retraits d'un régime de pensions ou d'un REER[93]. Tous ces éléments sont pourtant considérés dans la déclaration de revenus.

---

91. Statistique Canada, *Guide de l'utilisateur des microdonnées de l'Enquête sur la dynamique du travail et du revenu*, cat. nº 75M0001GPF, 1994, p. 147.
92. Statistique Canada (1994), op. cit., p. 144. Notons que cette substitution est également devenue possible avec le questionnaire long du recensement, puisque la question 51 du formulaire de 2006 offrait, pour la première fois, le même choix.
93. Régime enregistré d'épargne-retraite.

En deuxième lieu, le formulaire long du recensement fait place à la sous-catégorie des « autres revenus en espèces » (à l'intérieur de la catégorie des autres revenus) et l'illustre à l'aide des trois possibilités suivantes : les pensions alimentaires, les allocations de soutien des enfants et les bourses d'études. Cette liste n'est toutefois pas limitative. Le *Dictionnaire du recensement* ajoute trois autres exemples : le soutien périodique reçu de personnes ne faisant pas partie du ménage, le revenu reçu de l'étranger (sauf les dividendes et intérêts) et les indemnités de fin d'emploi et redevances[94]. Certains de ces éléments sont non imposables[95].

Tous ces exemples d'autres revenus en espèces n'épuisent pas le sujet. En effet, l'Enquête sur les dépenses des ménages (EDM) fait état de quatre types de revenus en espèces non énumérés dans le formulaire du recensement, mais que ce dernier devrait accueillir si un répondant rapportait l'un d'eux. Il y a d'abord les dons en argent offerts par des personnes n'appartenant pas au ménage (déjà mentionnés au paragraphe précédent). Mais s'y ajoutent les héritages en espèces, les règlements d'assurance vie et les gains provenant de jeux du hasard[96], tous non imposables.

Bref, on en retient donc que le dénominateur du rapport d'abordabilité, soit le revenu total avant impôt, est susceptible de varier quelque peu selon l'instrument à partir duquel il est établi. S'il est établi à partir de l'EDTR (comme il le sera dans cette recherche-ci), il sera influencé par la méthode fiscale et tiendra compte des gains ou des pertes en capital, et des retraits de régime de retraite ou de REER, mais ne pourra tenir compte des éléments non imposables des autres revenus en espèces.

À l'inverse, si le rapport est établi à partir des données du recensement (à l'instar de la série socioéconomique de la SCHL), il prendra en considération tous les autres revenus en espèces que les répondants auront bien voulu déclarer, mais omettra les gains et les pertes en capital et les retraits de régimes de retraite.

Pour clore ces éclaircissements, ajoutons un élément concernant le numérateur du rapport d'abordabilité – les frais d'habitation – pour y mentionner une distinction qui existe entre les données de l'EDTR et celles du questionnaire long du recensement. Dans l'établissement des frais d'habitation, et contrairement au questionnaire du recensement, l'EDTR ne fait pas de prélèvement d'informations auprès des répondants en ce qui a trait au coût de l'électricité, des combustibles et de l'eau, mais leur impute les données d'enregistrements similaires recueillies lors du recensement[97].

En ce qui concerne cette recherche, celle-ci travaille avec les données de l'EDTR. Il est donc acquis qu'elle utilise, dans l'établissement des rapports d'abordabilité, une version du revenu total qui colle davantage à la déclaration de revenus, et qui, de surcroît, est peu susceptible de recueillir les montants non imposables des autres revenus en espèces.

---

94. Statistique Canada, *Dictionnaire du recensement de 2001*, mars 2002, nº 92-378-XIF01000, p. 124 et 125.
95. Depuis 2002, les pensions alimentaires auraient tendance à ne pas être imposables. Le soutien périodique de personnes ne faisant pas partie du ménage ne l'est pas.
96. On retrouve ces éléments dans J. Luffman (2006), p. 18 et 23. L'Enquête sur les dépenses des ménages s'enquiert spécifiquement de ces éléments. Dans l'étude de Luffman, on utilise un modèle qui comporte une variable explicative du nom d'autres rentrées en espèces pour indiquer la présence de ceux-ci (voir le tableau 4 de l'article, p. 24).
97. EDTR, *Dictionnaire électronique des données*, variables sur le logement UECY25, UHTY25 et UWAY25, et W. Rea et J. Yuen, *op. cit.*, diapo 6.

ANNEXE 2

Comparaison des taux d'effort des ménages exploitant une ferme sur leur propriété avec celui des autres ménages, Québec, 2003, données pondérées complètes[1]

Taux d'effort	Exploitent une ferme sur leur propriété					
	Oui		Non		Total	
	n	%	n	%	n	%
Moins de 30 %	43,1	81,3	4 731,6	80,1	4 774,7	80,2
30 % et plus	9,9	18,7	1 172,4	19,9	1 182,3	19,8
Total	53,0	100,0	5 904,0	100,0	5 957,0	100,0

p χ^2 SUDAAN = 0,99

1 : Données complètes : pas de restrictions ou de seuils sur les enregistrements.

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

ANNEXE 3

Taux d'effort des ménages selon leur composition, Québec, 2003, données pondérées

Taux d'effort	Composition du ménage							
			Deux personnes ou plus					
	Une personne		Une famille économique		Deux familles économiques ou plus		Total	
	n	%	n	%	n	%	n	%
Moins de 30 %	1 148,7	65,8	3 435,3	88,8	183,6	79,0	**4 767,6**	81,5
30 % à 49,9 %	365,4	20,9	319,4	8,3	35,0	15,1	**719,7**	12,3
50 % et plus	232,7	13,3	112,5	2,9	13,9	6,0	**359,0**	6,1
Total	**1 746,7**	**100,0**	**3 867,1**	**100,0**	**232,4**	**100,0**	**5 846,3**	**100,0**

p χ^2 SUDAAN = 0,000

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

ANNEXE 4

Taux d'effort des ménages selon l'indicateur « Autres revenus », Québec, 2003, données pondérées

	Autres revenus					
	Oui		Non		Total	
Taux d'effort	n	%	n	%	n	%
Moins de 30 %	897,4	82,6	3 870,2	81,3	**4 767,6**	81,5
30 % à 49,9 %	134,4	12,4	585,3	12,3	**719,7**	12,3
50 % et plus	54,5	5,0	304,6	6,4	**359,0**	6,1
Total	**1 086,3**	**100,0**	**4 760,1**	**100,0**	**5 846,3**	**100,0**

p χ^2 SUDAAN = 0,602

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

ANNEXE 5

Taux d'effort des ménages selon le panel de l'enquête auquel appartient le soutien économique principal, Québec, 2003, données pondérées

Taux d'effort	Panel				Total	
	nº 3		nº 4			
	n	%	n	%	n	%
Moins de 30 %	2 258,0	81,5	2 509,6	81,6	**4 767,6**	81,5
30 % à 49,9 %	355,9	12,9	363,8	11,8	**719,7**	12,3
50 % et plus	155,3	5,6	203,8	6,6	**359,0**	6,1
Total	**2 769,1**	**100,0**	**3 077,2**	**100,0**	**5 846,3**	**100,0**

p χ^2 SUDAAN = 0,450

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

ANNEXE 6

Répartition des niveaux de revenu par principale source de revenus du ménage, fichier transversal, Québec, 2003, données pondérées

Niveaux de revenu total du ménage avant impôt	Principale source de revenus du ménage									
	Salaires et traitements		Revenu d'un travail autonome		Transferts gouvernementaux, autres revenus		Pensions de retraite, placements		Total	
	n	%	n	%	n	%	n	%	n	%
6 500 $ à 14 999 $	60,2	1,7	34,2	13,8	480,5	32,2	10,4	1,9	585,3	10,0
15 000 $ à 24 999 $	248,0	7,0	36,0	14,5	465,9	31,2	37,3	6,8	787,2	13,5
25 000 $ à 34 999 $	407,7	11,5	26,7	10,8	344,1	23,0	92,3	16,9	870,8	14,9
35 000 $ et plus	2 841,5	79,9	151,7	61,0	203,9	13,6	405,9	74,3	3 603,0	61,6
Total	3 557,4	100,0	248,6	100,0	1 494,3	100,0	545,9	100,0	5 846,3	100,0

p χ^2 SUDAAN = 0,0000

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier transversal 2003.
Compilation : Société d'habitation du Québec.

ANNEXE 7

Variables, niveaux et années pour lesquels le rapport de cotes du logit I est sensiblement plus accentué que celui du logit II, ou est seul significatif.

Variable	Niveau	Année
Âge	16 à 29 ans	2002
Âge	30 à 44 ans	2002
Âge	45 à 64 ans	2002
Âge	45 à 64 ans	2003
Âge	45 à 64 ans	2004
Princ. source de revenus	Travail autonome	2002
Princ. source de revenus	Transferts gouvernementaux, autres revenus	2002
Princ. source de revenus	Travail autonome	2004
Genre de famille	Personne seule	2002
Genre de famille	Personne seule	2003
Genre de famille	Personne seule	2004
Activité principale	Va à l'école	2002
Activité principale	Va à l'école	2003
Activité principale	Va à l'école	2004
Taille région résidence	Urbaine : 0 – 29 999 habitants	2002
Taille région résidence	Urbaine : 0 – 29 999 habitants	2003
Statut de déménagement	Déménagement	2004
Statut de déménagement	Déménagement après une séparation	2003
Statut de déménagement	Déménagement après une séparation	2004
Plus haut niveau d'études	Diplôme ou certificat universitaire	2004

ANNEXE 8

Extrait d'une présentation des modèles de trajectoires[98]

Modèles de trajectoires

Ce type de modèle fait partie d'une famille de modèles en statistique que l'on nomme **modèles de mélange fini** *(Finite Mixture Models)*.

Les **modèles de mélange fini** sont des modèles pour lesquels la fonction de densité est la somme pondérée de plusieurs densités, résultat de la modélisation d'une variable latente catégorielle et de la distribution conditionnelle des variables observées. La variable latente se révèle par la corrélation des variables observées.

Dans le cas de **modèles de trajectoires**, nous supposons qu'il y a hétérogénéité du comportement dans le temps de la variable d'intérêt. C'est cette caractéristique d'hétérogénéité qui nous amène à choisir ce type de modèle plutôt que le modèle de courbe de croissance traditionnel. Cette hétérogénéité est prise en compte par la modélisation de la variable latente catégorielle. Les modalités de la variable latente représentent des groupes pour lesquels la modélisation longitudinale des variables observées est homogène. Conditionnellement à ces groupes, les variables observées sont indépendantes.

Modèle théorique

La fonction de vraisemblance des modèles de trajectoires s'exprime comme suit :

$$\prod_i f(Y_i=y_i|Z_i=z_i, X_i=x_i)=\prod_i(\sum_k \pi_k(z_i, x_i)h_k(y_i|x_i))$$

où

π_k *()* représente la fonction de mélange de probabilités étant donné des covariables z_i et x_i. Cette fonction découle de la modélisation de la variable latente catégorielle. Cette fonction est une régression logistique binomiale ou une régression logistique multinomiale dépendamment si la variable latente a deux modalités ou plus de deux modalités. Les modalités de la variable latente sont représentées algébriquement par l'indice k et, dans le cas des trajectoires,

k représente le groupe.

$h_k()$ représente la distribution conditionnelle des variables observées y_i étant donné des covariables x_i et le mélange k.

98. *Modèles de trajectoires*, présentation de Danielle Forest du 28 mars 2007 à l'Université Laval (7 p.). Mme Forest est analyste au Centre interuniversitaire québécois de statistiques sociales (CIQSS), antenne de l'Université de Montréal.

ANNEXE 9

Taux d'effort annuel des ménages québécois, fichier longitudinal 2002-2004, données pondérées

Taux d'effort	2002		2003		2004	
	n	%	n	%	n	%
0 % à 29,9 %	4 044	83,7	4 072	84,3	4 027	83,3
30 % à 49,9 %	541	11,2	527	10,9	561	11,6
50 % et plus	247	5,1	233	4,8	244	5,0

Source : Statistique Canada, Enquête sur la dynamique du travail et du revenu, fichier longitudinal 2002-2004.
Compilation : Société d'habitation du Québec.

BÂTISSONS
DU MIEUX-
VIVRE
www.habitation.gouv.qc.ca

INTRODUCTION ANALYSES TRANSVERSALES DONNÉES LONGITUDINALES MODÈLES LONGITUDINAUX CONCLUSION 6. ANNEXES

Société
d'habitation
Québec